119

Reihe Ökonomie

Economics Series

Stigma and Social Control

Lawrence Blume

119
Reihe Ökonomie
Economics Series

Stigma and Social Control

Lawrence Blume

July 2002

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Lawrence Blume
Department of Economics
Uris Hall
Cornell University
Ithaca, New York 14850, USA
email: lb19@cornell.edu

Abstract

Social interactions provide a set of incentives for regulating individual behavior. Chief among these is *stigma*, the status loss and discrimination that results from the display of stigmatized attributes or behaviors. The stigmatization of behavior is the enforcement mechanism behind social norms. This paper models the incentive effects of stigmatization in the context of undertaking criminal acts. Stigma is a flow cost of uncertain duration which varies negatively with the number of stigmatized individuals. Criminal opportunities arrive randomly and an equilibrium model describes the conditions under which each individual chooses the behavior that, if detected, is stigmatized. The comparative static analysis of stigma costs differs from that of conventional penalties. One surprising result with important policy implications is that stigma costs of long duration will lead to increased crime rates.

Keywords

Crime, stigma, social norms

JEL Classifications

C730, Z130

Comments

I am grateful to the Cowles Foundation for Economic Research, the Department of Economics at Yale University, and the Institute for Advanced Studies in Vienna for very productive visits. I am also grateful for financial support from the John D. and Catherine T. MacArthur Foundation, the Pew Charitable Trusts, and the National Science Foundation. This research is a consequence of several stimulating conversations with Daniel Nagin, and has been helped along by comments from Buz Brock, Tim Conley, Steven Durlauf, Tim Fedderson, Josef Hofbauer, Karla Hoff, Chuck Manski, and Debraj Ray.

Contents

"…, let her cover the mark as she will, the pang of it will be always in her heart."

The Scarlet Letter
Nathaniel Hawthorne

1 Introduction

Consciously or not, in our interactions with others we identify markers that signal particular attributes. While the attribution signalled by a marker may be the result of rational inference, markers may also signal attributes by triggering stereotypes. Language usage, skin color, gender and occupation may all prompt individuals to infer that others possess a variety of behaviors and attitudes that have nothing at all to do with the markers they display. Markers that trigger negative stereotypes *stigmatize* their bearers.

The sociology and social psychology literature on stigma is replete with characterizations of the stigmatic process. Link and Phelan's (2001, p. 367) definition is particularly useful here because it addresses the social as well as the psychological aspects of stigma. They characterize stigma in terms of four interrelated components:

> In the first component, people distinguish and label human differences. In the second, dominant cultural beliefs link labeled persons to undesirable characteristics—to negative stereotypes. In the third, labeled persons are placed in distinct categories so as to accomplish some degree of separation of "us" from "them". In the forth, labeled persons experience status loss and discrimination that leads to unequal outcomes.

Stigma has both micro- and macrosocial consequences. Link and Phelan (2001) assert that research on stigma has attended primarily to its perception by individuals and its consequences for micro-level interactions. On the other hand, stigmatic markers classify entire groups of individuals. The aggregate of behaviors which respond to stigmatic markers has systemic implications for aggregate social and economic performance.

Both individual characteristics and individual behaviors are available for stigmatic representation. The stigmatization of race, gender, physical disabilities, mental illness and other characteristics is morally repugnant, and a continuing source of social ills. But the stigmatization of behaviors is the central mechanism for enforcing social norms. Stigma is thus essential to the production of what some call "social capital". Stigma enforces social norms by stigmatizing non-normative behavior. Here this mechanism is modeled in order to draw some conclusions about its efficacy. A concrete instance of the social control process modeled here is the stigmatization of certain kinds of criminals. Small-town newspapers routinely publish the names of those arrested (and not yet convicted) for driving under the influence of alcohol. Fears of public exposure are a strong incentive for tax compliance in some communities. But if everyone cheated or everyone drank, the social costs of exposure would be small.[1]

The stigma mechanism suggests a coordination game, with high- and low-activity equilibria. In the high-activity equilibrium, many people cheat, and so the stigma costs of cheating are small — cheating is not stigmatized. The low-activity equilibria has few cheaters and high stigma costs — cheaters are stigmatized. But my interest here is in the dynamics of stigma costs rather than in the description of static coordination games. The costs of being stigmatized are born in the future as well as today. Accounting for the future requires the consideration of stigma cost dynamics.

The model presented here is a dynamic population game model, loosely in the spirit of Blume (1993), Kandori, Mailath, and Rob (1993) and Young (1993). But it departs from these earlier dynamic models in its rejection of myopia. The usual population model from evolutionary game theory superimposes on a static game some dynamics meant to describe the flow of the distribution of strategies among the population. The chief drawback to this conventional modelling strategy is that individuals' decisions are unconnected to considerations of the future. This is frequently justified by claiming that individuals are myopic. With infinite subjective rates of time preference they have no need to consider the future consequences of their acts. The technical innovation of this paper is developing a population model in which individuals care about the future and account for the future evolution of the population in their decisionmaking. A suitable equilibrium concept is introduced and proved to exist. The comparative dynamics of equilibrium with respect to parameters of the model is worked out. The long-run implications for levels of criminal activity are demonstrated. Finally, a particular example is worked out in some detail, which demonstrates some additional consequences of equilibrium both in the short and in the long run.

2 The Model

Consider the criminal story — the stigmatization of tax cheaters and drunk drivers. The model contains two types of individuals: "Tagged" individuals have been caught and labeled as criminals at some point in the past. They have been stigmatized by having been caught displaying antisocial behavior. "Untagged" individuals bear no such label. There is a cost associated with criminal status. The magnitude of this cost is increasing in the size of the untagged population. There is no further stigma cost to additional crime once an individual is tagged. Tagged individuals revert to untagged status at random moments. The stigma of being tagged eventually wears off. This distinguishes crime, occupational choice and other stigmata marked by actions from those marked by characteristic, such as race, gender and mental illness, which may never wear off. Time is continuous and individuals maximize the present discounted value of a utility stream whose magnitude depends upon the following parameters.

Notation:

N The population size.

m_t The fraction of "other" untagged individuals.

p The arrival rate of criminal opportunities.

u The (random) utility reward for successfully completing a crime.

v The utility penalty for being apprehended and convicted. $\delta = u - qv$ is the expected instantaneous return to committing a crime.

$F(u)$ The cdf of the reward u.

q The probability of being captured and convicted after committing a crime. (We suppose the corresponding probability conditional on not committing a crime is 0.)

$c(m)$ The stigma (flow) cost of being tagged when the fraction of others who are untagged is m. This function is non-decreasing.

g The arrival rate of untaggings.

r The individual's instantaneous rate of time preference.

Intuitively, the equilibrium process evolves as follows. At random moments events happen to individuals in the population. Events are either crime opportunities or the removal of a tag (if present). Crime opportunities arrive to a given individual at rate p. When a crime opportunity arrives a return u to committing the criminal act is drawn (independently) from the distribution F. The individual, knowing the state and the return to crime, must decide whether or not to act. If she does not act, she receives 0 return and is not tagged. If she commits the crime and is not caught she gets the reward u and remains untagged. This happens with probability $1 - q$. But with the complementary probability q she is caught, pays a penalty v, and is tagged. Opportunities to become untagged arrive to a given individual at rate g. If she is not tagged, nothing happens, but if she is tagged, she becomes untagged.

These processes are independent across individuals. Given decision rules for the other individuals and the arrival rates of events described above, the process $\{m_t\}_{t=0}^{\infty}$ describing the evolution of states for individual i's decision process can be constructed.[2]

Individuals have beliefs about the state process. At a criminal opportunity, they act so as to maximize the expected present discounted value of their utility stream. An equilibrium tag process is the tag process which results when individuals' beliefs are correct.

3 Individual Choice and Social Equilibrium

The goal of this section is to define formally the tag process and the equilibrium concept just described. We will prove that there exists a unique symmetric equilibrium in state-dependent strategies. This equilibrium will be *monotone* in the sense that the probability of an untagged individual committing a crime is increasing in the state. Furthermore, the equilibrium strategies will involve randomization in at most one state.

This paper is concerned only with symmetric equilibria, in which all individuals adopt the same strategy, which depends only upon the state. Our definitions will be phrased accordingly, although generalizations of the definition (not the theorems) are obvious.

3.1 States and Strategies

At a criminal opportunity, each individual's decision will depend upon the immediate value of the crime and the state of everyone else. That state is an element of the decision maker's *state space* $\Omega = \{0, 1/(N-1), \ldots, 1\}$, with typical element m. Denote by m^+ the state $m + 1/(N-1)$, and by m^- the state $m - 1/(N-1)$.

At any moment of time, an individual can be in one of two conditions or *types*, untagged or tagged. A pure strategy for an individual currently of type $d = \{U, T\}$ (untagged, tagged) is a map which assigns to each state m and each immediate reward u a probability of committing the crime. Formally, a pure strategy for an individual currently of type $d \in \{U, T\}$ (untagged, tagged) is a map

$$\sigma_d : \Omega \times \mathbf{R} \to \{\text{Commit}, \text{Not}\}.$$

A (behavior) strategy for each type $d = \{U, T\}$ (untagged, tagged) is a map

$$\sigma_d : \Omega \times \mathbf{R} \to [0, 1],$$

where $\sigma_d(m, u)$ is the probability that a type d individual commits a crime with reward u in population state m. A *strategy* is a pair of type strategies

$\sigma = (\sigma_U, \sigma_T)$. A strategy is *monotonic* if, for each type, the probability of committing a crime falls with the state. That is, the greater the fraction of untagged individuals, the lower the probability of anyone's committing a crime. A *reservation strategy* exhibits a return threshold u^* above which crimes are committed and below which they are not.

Definition 1. *A strategy σ is* monotonic *if for all d and u, $\sigma_d(m, u)$ is non-decreasing in m. It is a* reservation strategy *if for all d and m there is a u^* such that $\sigma_d(m, u) = 1$ for $u > u^*$ and $\sigma_d(m, u)$ is 0 for $u < u^*$.*

At any instant of time at most one individual has an opportunity of some kind, and so the value of the state process can change by at most $\pm 1/(N-1)$. Consequently the state process $\{m_t\}_{t \geq 0}$ is a *birth-death process.* The birth and death rates are determined by the (mixed) strategy σ. Suppose the population state is m. A "birth" occurs when a tagged individual becomes untagged. The rate at which opportunities arrive to tagged individuals is $(N-1)(1-m)$, and the probability that an event is an untagging is g, so the birth rate is

$$\lambda_m = (N-1)(1-m)g \tag{1}$$

A death occurs when an untagged individual commits a crime and is caught. The death rate for individual i's state process depends upon her type. Define for each type d and state m $\sigma_d(m) = \int \sigma_d(m, u) dF(u)$ to be the probability that a decisionmaker of type d will commit a crime in state m. We abuse notation by using this σ this way because it will be clear both from the context and by the number of arguments whether or not we want to condition on the return. Suppose she is untagged and the state is m. Any other untagged individual also sees state m, and so the death rate is

$$\mu_m^U = (N-1)mp\sigma_U(m)q \tag{2}$$

If individual i is tagged and is in state m, then any other untagged individual sees state m^+. in this case the death rate in state m is

$$\mu_m^T = (N-1)m^+ p\sigma_U(m^+)q$$

3.2 The Individual's Decision Problem

When a criminal opportunity arrives, the individual who has received it must decide whether or not to commit a crime. The rational decisionmaker must account for the immediate expected return to a crime, and also for the stream of stigma costs. If the individual were fully rational and alert to all strategic interactions, then in computing the expected present discounted value of the stigma cost of a crime she would account for the effect of her own tagging on the propensity of others to commit crimes. When computing the evolution of states, she would assume death rates μ^T conditional on her being tagged, and μ^U conditional on her being untagged. We will assume that individuals are less strategic than this. They account for both instantaneous and dynamic effects, but they neglect the incentive effect of their own decisions on others. Whether tagged or not, they they assume the state process has the same death rates $\mu = \mu^U$.[3]

The individual's decision problem can be formulated as a dynamic program. The program is described by three independent processes: The arrival process for criminal opportunities, the arrival process for untaggings, and the state process. The *criminal opportunity process* is a rate-p Poisson process and the *untagging process* is a rate-g Poisson process. The *state process* has the birth and death rates λ_m and μ_m just described. All these rates can be derived from the parameters and σ, the strategy employed by others. Thus given the parameters, each individual's decision problem is characterized by (σ_T, σ_U).

At a decision opportunity the individual knows the history of the tag process, her individual history, and the payoff u to the current crime . She chooses an action, to commit a crime or not, so as to maximize the expected present value of her utility stream. Her instantaneous utility depends upon the state of the population process, her current type, and whether or not she has a decision opportunity.

At a moment which is not a decision opportunity she is either tagged or untagged. If untagged, her instantaneous utility is 0. If tagged, it is $-c(m_t)$. At a decision opportunity she is either tagged or untagged. She decides whether or not to commit a crime. If she chooses not to commit a crime, her instantaneous utility is that which she would receive were she not to have a decision opportunity. If she commits a crime, her utility depends upon whether or not she is caught, and whether or not she is tagged. If she is already tagged, she receives reward u. She pays a penalty v if caught, so her net return if caught is $u-v$. She has an instantaneous stigma cost flow of

$c(m_t)$ which is independent of her decision and whether or not she is caught.

If she is untagged, she has the same immediate net reward structure: u for the crime, less a penalty v if she is caught. But if she is caught her status switches from untagged to tagged, and so she begins to pay the stigma cost flow $c(m_t)$ which she did not bear previously. This flow lasts a random amount of time, independent of her future decisions, until the tag disappears.

Consider a sample path from the meet of the three processes and a strategy. The strategy generates a stream of utility. The value of the strategy on that path is the present discounted value of the utility stream (discounted at rate r), and the expected value of a strategy is the expectation of this value over all sample paths of the meet. An optimal strategy maximizes expected value over all strategies.

3.3 Equilibrium — Existence and Monotone Comparative Statics

We will be looking for symmetric equilibria; that is, equilibria in which all individuals use the same strategy. We can already see this in the construction of the individual decision problem. In principle the definition of the problem can be extended to encompass different individuals using different strategies. At that point, however, the birth-death formalism is lost because in order to keep track of the evolution of m_t we would need to know the identities of the tagged individuals.

Definition 2. *A strategy $\sigma = (\sigma_U, \sigma_T)$ is a* population equilibrium strategy profile *if σ is optimal for the individual decision problem with parameters σ.*

This equilibrium is not Nash! It fails to be Nash because, as we discussed earlier, each individual neglects the impact of her policy on the evolution of other individuals' decisions. Nonetheless, this equilibrium concept is close to Nash, and in some appropriate large-numbers limit it would be Nash. To the extent that equilibrium fails to be Nash, it is a consequence of a small numbers problem.

It would seem that somehow a strategic complementarity must exist. As more people commit crimes, more people become tagged and the expected stigma cost of committing a crime falls, making crime more attractive. However, at this point we cannot even define a meaningful notion of strategic complementarity since there is no natural order for the strategy space with respect to which there might be increasing differences and the like. Nonetheless we will see that the strategic complementarity intuition is essentially correct.

The first Theorem describes population equilibria:

Theorem 1. *A population equilibrium exists, every population equilibrium uses monotonic reservation strategies and is pure in all but at most one state, and $\sigma_T(m, u)$ is 1 if $\delta > 0$ and 0 if $\delta < 0$.*

Existence is not a surprise. It is also not surprising that if crime does not pay in the short run, it never pays. Monotonicity for $\delta > 0$ is a consequence of the birth-death construction and the monotonicity of instantaneous rewards with respect to the state m_t. It is not hard to see that for an open and dense set of parameter values, equilibrium is pure. The case $\delta = 0$ is just like $\delta > 0$ for all states except $m = 0$, where anything can be chosen.

It will prove useful to track the states at which, for a given utility level, crime begins to take place with positive probability.

Definition 3. *Let σ denote an equilibrium strategy. For each utility level u, let*

$$m_u = \max\big\{\{m : \sigma_U(m, u) > 0\} \cup \{+\infty\}\big\}.$$

The state m_u is the switch point *for u in strategy σ.*

The presence of strategic complementarities has implications for the dependence of equilibrium on model parameters. Say that strategy σ is "as criminal as" strategy σ' if, in every event and for every type and payoff, the probability of committing a crime under σ is at least that under σ'.

Definition 4. *Strategy σ is* at least as criminal as *strategy σ' (write $\sigma \succeq \sigma'$) iff for each state m, type d and payoff u, $\sigma_d(m, u) \geq \sigma_d(m', u)$.*

Real-valued parameters are ordered in the usual way. Payoff distributions are ordered by means. That is, $F \geq_m G$ iff the mean of F is at least as big as the mean of G. Cost functions are ordered pointwise. $c(\cdot) \geq d(\cdot)$ if for all m, $c(m) \geq d(m)$. With respect to these orderings and the "at least as criminal as" ordering there is a comparative equilibrium result.

Theorem 2. *For each vector of parameter values the set of equilibria are totally ordered by $\succeq$, and there is a greatest and a least equilibria. These extreme equilibria are increasingly criminal in F and r, and decreasingly criminal in c and v.*

The effects of changes in g and q are ambiguous, and will be discussed further below.

3.4 A Discrete Example

The simplest examples to compute have a distribution of utilities with two possible values: One such that the crime never pays and one such that crime sometimes pays. Suppose that a cost function is defined on the interval $[0, 1]$. It is non-decreasing and piecewise-continuous. There are two possible reward values: u_h, realized with probability $\epsilon > 0$, very small, which is so high that any such criminal opportunity is always acted upon, and u_l which may or may not be acted upon by untagged individuals, depending upon the expected present discounted value of stigma costs.

Equilibrium is characterized by the switch point s, the largest state in which a crime with reward u_l will be carried out by an untagged individual, and p_s, the probability of carrying out a crime in state s. For all states less than s, the probability of committing a crime with reward u_l is 1, and p_s will be 1 as well, unless the individual is indifferent between committing the crime and abstaining.

Equilibria in this model are easily computed by methods discussed in section 6. This makes possible an investigation of the comparative statics of changes in parameters g and q which were not determined by complementarity arguments.

The effects of changes in g and q are ambiguous. On the one hand, increasing q increases the probability of being caught and paying direct and stigma costs. As in Becker's (1968) model of criminal deterrence, this *neoclassical effect* reduces criminal activity. But a second consequence of increasing q is a *social interaction effect*. Increasing q increases the number of tagged individuals, thereby reducing the stigma costs of having been caught committing a crime. This effect increases criminal activity. The comparative statics of a change in q depends upon the balance between these two effects. A downward sloping relationship between the probability of getting caught q and criminal activity is easily illustrated in simple examples.

In the following example, $v = 0$ to focus on stigma costs. A slice of the equilibrium correspondence, plotting the relation between equilibrium s and q.[4] Every dot indicates an equilibrium switchpoint. This example

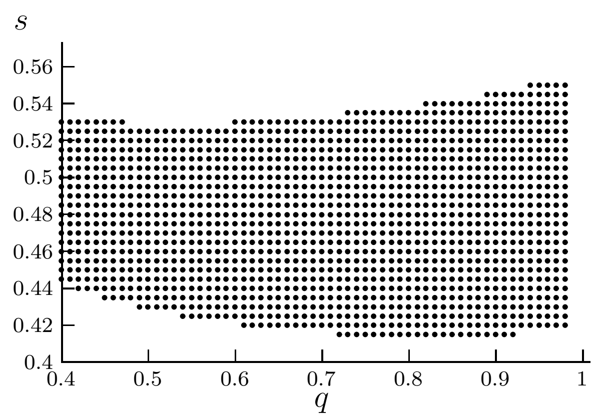

Figure 1: The Equilibrium Correspondence: s vs. q.

illustrates nicely the possibility of multiple equilibria. The example exhibits the neoclassical relationship between the probability of getting caught and the criminality of the equilibrium strategy over most of its range. But for q large enough, the social interaction effect dominates. The comparative statics changes direction; increasing q increases the criminality of the equilibrium strategy set.

Not surprisingly, similar effects are at work in the relationship between equilibria and the parameter g, the arrival rate of untaggings. When

g increases, the conventional effect is that stigmatic markers are held for a shorter period of time, and so stigma costs decrease. Hence criminal activity should become more likely. The social interaction effect, however, has it that fewer individuals are tagged at any moment in time, and so the flow costs of stigma have increased. The social interaction effect dominates in the downward-sloping part of the following figure.[5] Expected duration of the

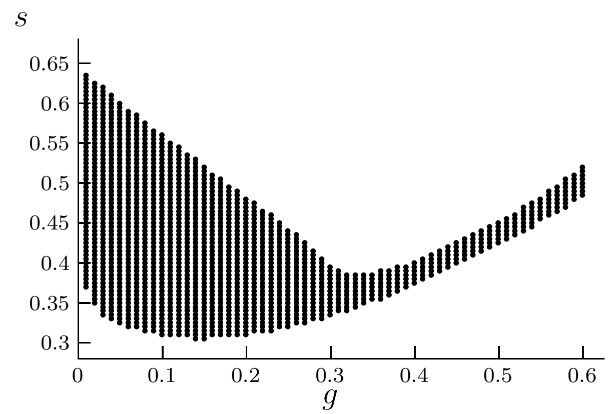

Figure 2: The Equilibrium Correspondence: s vs. g.

stigmatic punishment, $1/g$, is in fact a policy variable of those courts which administer so-called “shaming-penalties”.[6] Book (1999) relates that in colonial Williamsburg a convicted thief was nailed by his ear to the pillory. At the completion of his sentence his gaolers ripped him from the pillory without first removing the nail, thereby “earmarking” him for life. That is, $g = 0$. Overly long punishments can have the perverse effect of reducing the incentives to avoid criminal activity. The positive effect on the cost of increasing the waiting time τ is more than offset by the decreases in cost created by changes in the $\{m_t\}_{t\geq 0}$ process.

4 The Equilibrium Tagging Process

The ebb and flow of criminal behavior, long run averages and likely short run paths, are properties of the population tag process. This process is derived from the equilibrium strategy and the assumptions about the stochastic processes generating crime opportunities, capture, and untagging.

4.1 Birth and Death Rates

Define the population state space $S = \{0, 1/N, \ldots, 1\}$. Let $\{n_t\}_{t\geq 0}$ denote the *population process* for a population of size N. This is a birth-death process with birth rates κ_n^N and death rates ν_n^N:

$$\begin{aligned} \kappa_n^N &= N(1-n)g \\ \nu_n^N &= Nnp\sigma_U\Big(\frac{Nn-1}{N-1}\Big)q \end{aligned}$$

The argument of σ_U is computed as follows. In population state n, an untagged individual sees $Nn-1$ other untagged individuals, so the fraction of others who are tagged is $m(n) = (Nn-1)/N$. These birth and death rates completely characterize the population process, including its short run and long run behavior. Let n^+ and n^- denote the next biggest and next smallest states to n, $n + (1/N)$ and $n - (1/N)$, respectively.

4.2 Equilibrium in the Long Run

For any equilibrium, the long run behavior of the population process is described by its invariant distribution. The state process is always ergodic. Since the birth rate is strictly positive for any $n > 0$, it is possible to reach 1 from any state. Consequently the state process has a unique invariant distribution. For instance, if there is a state m^* such that no crimes are committed for $m > m^*$, then the unique invariant distribution puts all its mass on the state $n = 1$.

One virtue of the birth-death formalism is that the invariant distribution is easily computed. The invariant distribution π is characterized by the relationship

$$\pi(n)\kappa_n^N = \pi(n^+)\nu_{n^+}^N$$

Iterating this relationship, we have that if there is a state n' such that $\nu^N_{n'} = 0$, then $\pi(n) = 0$ for all $n < n'$, and if $n' = \max\{n : \nu^N_n = 0\}$, then for all $n > n'$,

$$\begin{aligned}\frac{\pi(n)}{\pi(n')} &= \prod_{k=n'}^{n^-} \frac{\kappa^N_k}{\nu^N_{n^+}} \\ &= \prod_{k=n'}^{n^-} \frac{1-k}{k} \frac{g}{pq\sigma_U\left(\frac{Nk-1}{N-1}\right)}\end{aligned} \tag{3}$$

where the index is understood as being incremented in units of size $1/N$.

The invariant distribution allows for the calculation of such variables as the long run crime rate. If the population is in state n, the rate at which individuals commit crimes is

$$\eta(n) = N(1-n)p + Nnp\sigma\left(\frac{Nn-1}{N-1}\right);$$

The $N(1-n)$ tagged individuals commit crimes at rate p; every crime that comes their way. The Nn untagged individuals commit crimes at the lower rate given by their strategy. The average of this function with respect to the invariant distribution gives the long run crime rate.

The comparative equilibrium analysis of Theorem 2 has straightforward implications for the long run behavior of the population state process. Let x be a parameter with respect to which equilibrium is increasing. As x increases, the death rates in each state increase. The theorem states that in each state $m = (N - Nn)/(N - 1)$, the probability of committing a crime is (weakly) increasing, and so, according to the definition, is the death rate ν^N_n. The birth rates are fixed by the parameter g, and so remain unchanged. Comparative static analysis of equilibrium strategies has straightforward implications for changes in death rates, and therefore for changes in the invariant distribution.

Theorem 3. *The invariant distribution is non-increasing with respect to parameters F and r in the sense of first-order stochastic dominance, and non-decreasing in c and v.*

Corollary 1. *The crime rate $\eta(n)$ is non-decreasing with F and r, and non-increasing in c and v.*

4.3 Tag Duration in the Long Run

The failure of the conventional intuition to predict the effects of changes in g and q on the criminality of equilibrium strategies depends upon the values of other parameters. Cases other than those presented in the previous section present the predicted comparative statics. But the effects of these parameters is due not just to the equilibrium strategies but also to the rate at which those committing crimes are tagged, and the rate at which they are subsequently untagged. Due to this second effect, the counterintuitive effect of increasing g on the invariant distribution and on the long-run crime rate is universal for large enough parameter values.

Denote by σ^* the map from $\Omega \times \mathbf{R} \to \{\text{Commit}, \text{Not}\}$ such that $\sigma(\omega, u) = \text{Commit}$ if and only if $u - qv \geq 0$. Recall that regardless of parameter values, $\sigma_T = \sigma^*$. Any strategy σ_U more criminal than σ^* is strongly dominated by it, since the additional opportunities σ_U accepts that σ^* does not have negative immediate rewards. More generally, if σ_d is any reservation strategy which in some recurrent state ω has a threshold $u < qv$, it is strictly dominated by the strategy which raises the threshold in state ω to qv.

Let $\eta^* = p\big(1 - F(qv)\big)$ denote the crime rate which would be observed if all untagged individuals acted according to σ^*. From these considerations the following lemma is obvious.

Lemma 1. *The long run equilibrium crime rate $\eta(n)$ is bounded above by η^*.*

We should expect this upper bound to be achieved for large values of g. When g is large, the expected duration of a tag is small, and status costs become negligible. The neoclassical effect should dominate the social interaction effect. This is true, but the bound can also be achieved for small g. Intuitively, when g is sufficiently small, any tag lasts a very long time. In the long run, most of the population will be tagged most of the time, and therefore most behavior is governed by $\sigma_T = \sigma^*$, rather than σ_U (whatever it may be). Let μ_0 denote the probability distribution on S which puts all its mass on 0, and let μ_1 denote point mass at 1.

Theorem 4. *Suppose that $F\big(qv + qc(1)/r\big) < 1$. Then $\lim_{g\to 0} \mu_g = \mu_0$. Suppose that $F(u)$ is continuous at $u = qv$. Then $\lim_{g\to\infty} \mu_g = \mu_1$. In both cases, $\lim_{g\to 0} \eta(n) = \eta^*$.*

If status costs ever have an impact on choice, then the long run crime rate must be decreasing in g over some range of values.

4.4 The Discrete Example: Short Run, Large N

For any differentiable function $f : S \to \mathbf{R}$ the birth and death rates give a differential equation which characterizes the evolution of the conditional expectation of f through time:

$$\begin{aligned}\frac{d}{d\tau}\mathrm{E}\{f(n_{t+\tau})|n_t = n\}\big|_{\tau=0} &= \kappa_n^N\Big(f(n^+) - f(n)\Big) + \nu_n^N\Big(f(n^-) - f(n)\Big) \\ &= \kappa_n^N f'(n)\frac{1}{N} - \nu_n^N f'(n)\frac{1}{N} + O(N^{-2}) \\ &= \Big((1-n)g - np\sigma_U\big(m(n)\big)q\Big)f'(n) + O(N^{-2})\end{aligned}$$

If $f(n) = n$, then

$$\frac{d}{d\tau}\mathrm{E}\{n_{t+\tau}|n_t = n\}\big|_{\tau=0} = (1-n)g - np\sigma_U\big(m(n)\big)q + O(N^{-2})$$

The differential equation

$$\dot{n} = (1-n)g - np\sigma_U\big(m(n)\big)q$$

is the *mean field equation* of the model. For the utility distributions considered in this and the next section, it is of the form

$$\dot{n} = (1-n)g - npq(1-\epsilon) \qquad \text{for } n < s, \tag{4a}$$

$$\dot{n} = (1-n)g - npq\epsilon \qquad \text{for } n > s, \tag{4b}$$

with steady states

$$n_l = \frac{g}{g + pq(1-\epsilon)} \quad \text{and} \quad n_h = \frac{g}{g + pq\epsilon}$$

(one for each branch), and solution

$$n(t) = n_l + \big(n(0) - n_l\big)e^{-\big(g+pq(1-\epsilon)\big)t} \qquad \text{for } n(0) < s, \tag{5a}$$

$$n(t) = n_h + \big(n(0) - n_h\big)e^{-(g+pq\epsilon)t} \qquad \text{for } n(0) > s, \qquad (5b)$$

The mean field equation characterizes the behavior of the process over finite time horizons in large populations. The following result is well-known in the literature on density-dependent population processes (Ethier and Kurtz 1986, Chapt. 10, Theorem 2.1).

Theorem 5. *Let $\{\sigma^N\}_{N=1}^{\infty}$ denote a sequence of equilibria in a population of size N, such that the switch points s^N converge to a limit s. Suppose that $n(t)$ is the solution to the differential equation (4) from initial condition $n_0 \neq s$, and let n_t^N denote the random variable that describes the population state at time t in a population of size N beginning from initial condition $n^N(0) = n_0$. Then for all $T > 0$, $\lim_{N\to\infty} \sup_{t\leq T} |n_t^N - n(t)| = 0$ a.s.*

The steady states are states in which the rate at which untagged individuals are tagged just equals the rate at which tagged individuals are untagged. The two different steady states correspond to the different rates at which untagged individuals commit crimes above and below the switch point. This is not to say that the two steady states are realized in practice. If for instance, $s > n_h$, then n_h is actually in the regime of low criminal activity. Starting from a high state, n_t will move downward according to equation (5a) until state s is reached, and then continue moving down through n_h according to equation (5b).

4.5 The Discrete Example: Long Run, Large N

In the discrete model with $\epsilon > 0$ the equilibrium odds ratios for an equilibrium with switch point m_u and no mixing are given by the following formulas. Let $n_u = m_u + (1 - m_u)/N$.

$$\begin{aligned}\frac{\pi(n)}{\pi(0)} &= \prod_{k=0}^{n^-} \frac{1-k}{k} \frac{g}{pq\sigma_U\left(\frac{Nk-1}{N-1}\right)} \\ &= \begin{cases} \binom{N}{Nn^-}\left(\frac{g}{pq(1-\epsilon)}\right)^{Nn^-} & \text{if } n^- < n_u, \\ \binom{N}{Nn^-}\left(\frac{g}{pq(1-\epsilon)}\right)^{Nn_u}\left(\frac{g}{pq\epsilon}\right)^{N(n^- - n_u)} & \text{if } n^- \geq n_u. \end{cases}\end{aligned}$$

A simple asymptotic approximation describes the shape of the invariant distribution for large populations. This analysis follows Blume and Durlauf (1998). Let

$$\rho(n) = \frac{1}{n^n(1-n)^{(1-n)}}\left(\frac{g}{pq\epsilon}\right)^{\min\{n,n_u\}}\left(\frac{g}{pq(1-\epsilon)}\right)^{\max\{0,n-n_u\}}$$

Use Stirling's formula to estimate the factorials and take limits:[7]

$$\pi^N(n) \approx \frac{z_N \rho(n)^N}{\sqrt{2\pi n(1-n)}}$$

where z_N is a normalizing factor. Thus the behavior of the invariant distribution for large N is governed by the behavior of $\rho(n)$. Taking derivatives of $\rho(n)$ with respect to n, we see that $\rho(n)$ has two local maxima n_l and n_h, with n_u in between. Raising ρ to the power n has the effect of making the function steeper without changing the location of the maxima. Thus the invariant distribution piles up at the steady states of the mean field approximation as N becomes large.

Typically there will be only one global maximum, and so as N becomes large, the invariant distribution will pile up on one and only one of the steady states. That is, the invariant distribution will converge with N to point mass on one of the steady states as N grows large. Which steady state gets the mass in the long run is determined by the location of the limit switch point. For example, the following plot shows invariant distributions over part of the range at two different switch points. The parameter values are $p = 0.15$, $g = 0.1$, $q = 0.6$, $r = 0.9$, $c = 2.5$, $\delta = 1.0$ and $\epsilon = 0.1$. With these values, the steady states are $n_l = 0.552$ and $n_s = 0.917$. The range of equilibrium switch points is $(0.752, 0.800)$. The plot shows the invariant distributions for switch points $n_u = 0.76$ and $n_u = 0.78$. They share the first piece in common, because, up to the rescaling coefficient z_N, the first piece of the distribution depends on n_u only for its stopping point. When $n_u = 0.76$ the switch comes earlier, and so the second hump is taller than the first. When $n_u = 0.78$ the switch is later, and the reverse is true. The invariant distribution tends to point mass at n_l in the latter case, and to point mass at n_s in the former case. Of course it is possible to choose parameter values where there is no transition. For instance, simply make ϵ small.

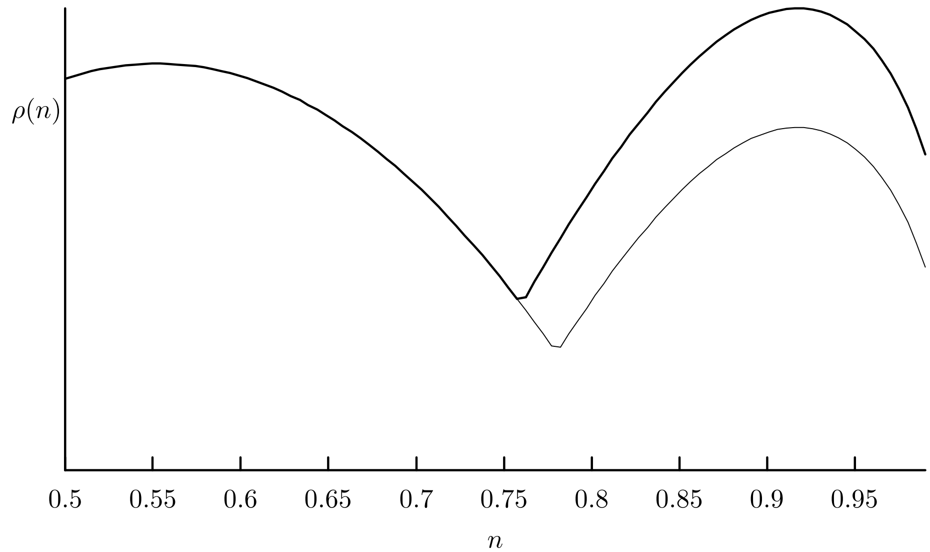

Figure 3: Invariant distributions for two different switch points.

5 Conclusion

A complete account of stigma as a social control process requires an analysis of how stigmatic attributions are formed. In fact there are many accounts. Closest to contemporary decision theory is a model of stereotyping which recognizes the use of overly simple models for categorization as an efficient allocation of cognitive resources.[8] I offer no such account here. Instead I make micro-level assumptions about the incentive effects of stigma. The driving assumption of my analysis is that the cost of being stigmatized, however it is realized, is low when many people bear the marker, and highest when only a few are so marked. Even this assumption would run afoul of some coherent theories of stigmatization.[9] To the extent that stereotyping is a statistical phenomenon, it is hard to form stereotypes if the incidence of the marker is low. Here I envision the social control process runing on a time scale which is short relative to the persistence of stereotypes, so that the stability of the stigmatic power of particular markers is not at issue. To the extent that the group at risk of stigmatization is large, the social cost of discipline may be too high when a large fraction of the group is tagged. The social cost of stigmatizing drunk drivers is small; the wage effects of a social custom not to hire those who park illegally could be enormous. Here I envision that the behaviors being stigmatized are seriously considered by or feasible to a small subset of the total population.

The efficacy of the stigma as a social control mechanism raises the question of if, and how, it can be used as a policy tool. Lessig (1995) offers a lovely anecdote about a government attempt to manipulate the stigma cost function c. In the late 1950's motorcycle helmets were beginning to leak from western Europe into the Soviet Union, which produced none. For the Soviet leadership, the medical benefits of wearing helmets were exceeded by the social cost of an invasion by a western style. "Thus began an extraordinary and self-conscious campaign by the Soviet government to vilify the wearers of motorcycle helmets. Cartoons appeared in the popular (read: government-controlled) press, mocking the 'white heads' on cycles. By the early 1960s, people began wearing helmets only at night, to avoid easy detection."[10] According to Lessig, helmets were never banned outright, suggesting that the stigmatization of riders was effective enough. Soon enough, however, the Soviets began to produce their own helmets, and with the availability of Soviet helmets, the campaign changed. Instead of stigmatizing helmet-wearers, it switched to stigmatizing those who imported helmets. The stigma cost of wearing helmets fell, and helmet usage increased.

Kahan (1997) claims a more subtle example of stigma manipulation in the policy of rewarding inner-city high school students who turn in peers carrying guns. In his account, not carrying a gun is a stigmatic marker. He argues that the reward policy succeeds because it manipulates stigma as well as having a direct effect on the stock of guns. The stigmatic effect is that when some students are out for the reward, displaying one's gun becomes more costly. If gun owners become more reluctant to display them, then the meaning of the marker changes, the stigma costs of not carrying falls, and the incentive to carry a gun is reduced.

A third, prominant example of promoting social control through stigma management is the increased use of shaming punishments in the United States and Great Britain.[11] In colonial America shaming could be for life. In modern times, shaming penalties are frequently seen as part of a probation sentence, a less costly and disruptive approach to behavior modification than prison for appropriate convicts.

As a tool of social control, stigma can be mismanaged. Theorem 4 shows that increasing the duration of stigma (decreasing g) will ultimately increase the long run crime rate. The largest possible crime rates are achieved when the duration is extremely long. This point is not merely of academic interest. Third strike drug offenders are banned for life from receiving a variety of federal benefits, including food stamps and temporary assistance to needy families available under the Personal Responsibility and Work Opportunity Reconciliation Act of 1996.[12] In some states, businesses requiring licenses cannot obtain them if they employ convicted felons, no matter how old the offense. This effective lifetime employment ban bars former convicts from working in, among other locations, barber shops and automobile body shops. In a similar vein, easy access to criminal records makes it easier for employers not obligated by law to nonetheless turn down applicants with criminal records. For instance, The Fair Credit Reporting Act prohibited the reporting of convictions more than seven years old, until this provision was deleted in 1998.[13] This information be be socially useful for its signal value, but Theorem 4 suggests that the availability of such old information may well have a negative deterrent effect on crime.

Another important counter-productive effect of stigmatization is not captured in this model. When "normal" society shuns the stigmatized, some may seek to shed the stigma (for instance, by having physical deformities corrected) or to overcome it by excelling in other dimensions. Others may respond by joining together with other stigmatized individuals to create "counter-communities" — communities in which the stigmatized activity is ignored or even becomes a source of status.[14] Newman (1999) writes about the stigma teenagers attached to "flipping burgers" in Harlem and other poor New York neighborhoods and, in particular, about the strategies young workers employ to defend themselves against the jokes and ridicule directed at them.

> ...it is clear that the workplace itself is a major force in the creation of a 'rebuttal culture' among these workers. Without this haven of the fellow-stigmatized, it would be very hard for urger barn employees to retain their dignity. With this support, however, they are able to hold their heads up, not by defininining

> themselves as separate from society, but by callking upon their commonality with the rest of the working world.[15]

Informal social control of deviance presumes a community which is sufficiently cohesive, well-organized, and has sufficient resources to enforce social norms. Elijah Anderson's (1990) ethnography of "Northton" describes the how the clash between "decent" norms (family life, hard work, church-going) and "streetwise" norms (associated with crime and the drug culture) is facilitated by a weakened structural fabric. The negative correlation of social organization and crime rates appears in empirical analyses as well. Sampson and Groves (1989) found in British data that neighborhoods with lower levels of social organization had higher levels of violent and property crimes. Unsupervised teen groups were the largest contributors to the violent crime rate, while local friendship networks and organizational participation had a large negative impact on robbery. Most surprisingly, the effect of the measured indices of social organization on crime exceeded the direct effects of socio-economic status. One source of disrupted friendship networks and broken families is the high incarceration of young male African-Americans. The legislative response to the crack epidemic of the 80's has been massive mandatory minumum sentences.[16]

A reinforcing effect of stigma not captured here is its effect on labelling the boundaries of normative behavior. When Hester Prynne is marked with the scarlet letter 'A', not only is she stigmatized, but the community reaffirms for itself the labelling of adulterous behavior as deviant. A contemporary (and perhaps non-fictional) instance of this labelling effect is the so-called "broken windows" theory which lies behind "order maintenance policing".[17] Lessig (1995), Kahan (1997) and others argue that the power of law to establish social boundaries and create categories for stigmatization is not yet fully appreciated as a source of social control.

The account of stigma and the enforcement of social norms presented here extends the evolutionary game theory paradigm by offering a richer account of individual choice. In particular, forward looking behavior is rarely studied, and almost never in stochastic models.[18] Stigma is in essence a dynamic phenomena. Its costs are born in the future, and the magnitude of those costs are determined by the future actions of others. This is why the pop rational actor social science accounts which, at their best, make reference to some kind of evolutionary game dynamics in a coordination game, seem so shallow. For instance, it would be hard to formulate a question about the effect of punishment duration in such models. The subject of evolutionary game theory is the dynamics of player choice. Recognizing players as intertemporal decision makers models opens up a variety of new modelling opportunities. Evolutionary game theory has been nearly devoid of serious applications to the social sciences. The premise of this paper is that deeper models of individual choice will provide evolutionary game theory with the wherewithal to address social issues.

6 Proofs

This section contains proofs of Theorems 1 through 4 and details on computing equilibria.

6.1 Theorems 1, 2 and 3

The proofs of Theorems 1 and 2 rely on the strategic complementarity that works through the stigma cost of crime. More criminal strategies lead to more tagged individuals, which lowers the stigma costs, thereby making crime more profitable. The complementarity appears twice in the proof: First to show that the optimal response to any strategy is a monotonic reservation strategy, and second to work the fixed point argument to get the existence of equilibrium results and to sign the dependence of the equilibrium strategy set on parameters.

In the individual's decision problem, *states* are the number of others who are tagged. The assumptions of the model implies that each individual takes the state process to be a birth-death process with some given rates. States evolve, and independently, *events* happen. An event is the arrival of either a criminal opportunity or an untagging. The event process is the superposition of two independent Poisson processes: A rate p process for criminal opportunities and a rate g process for untaggings. The event process is a rate $p + g$ Poisson process, and the probability that a given event is a criminal opportunity is $p/(p+g)$. Types of events are uncorrelated over time. See Kingman (1993) for details.

Construction of value functions and evaluation of policies involves comparing the values of functionals along paths of the state process. This is done with a coupling argument. Let ω denote a path of the birth-death process $\{m_t\}_{t=0}^{\infty}$, and define the function $f(\omega) = \int_0^{\infty} e^{-\lambda t} g(\omega_t)dt$ on paths.

Lemma 2. *If $\{m_t\}_{t=0}^{\infty}$ is a birth-death process and $g(m)$ is non-decreasing in m, Then $E\{f(\omega)|\omega_0 = m\}$ is non-increasing in m. If $g(m)$ is not constant, then the conditional expectation is strictly decreasing in m.*

Proof. Choose $m' < m''$, and construct the stochastic process $\{(x_t, y_t)\}_{t=0}^{\infty}$ with $(x_0, y_0) = (m', m'')$ as follows: Let $s = \inf\{t : x_t \geq y_t\}$ denote the *coupling time* of the x_t and y_t processes. Let x_t evolve according to the birth and death rates of the m_t-process. Let y_t evolve according to the rates for the m_t-birth-death process, independently so long as $x_t < y_t$, that is, so long as $t < s$. Observe to that almost surely $s < \infty$ and that $x_s = y_s$. Let $y_t = x_t$ for $t \geq s$. Observe that each marginal process is a birth-death process evolving according to the rates for the m_t-process. Furthermore, almost surely $x_t \leq y_t$ for all t.

Now consider the flows $g(x_t)$ and $g(y_t)$. Clearly $g(x_t) \leq g(y_t)$ almost surely for all t. Consequently for almost all (x_t, y_t) paths,

$$\int_0^{\infty} e^{-\lambda t} g(x_t)dt \leq \int_0^{\infty} e^{-\lambda t} g(y_t)dt$$

The expectation of the left hand side is $E\{f(\omega)|\omega_0 = m'\}$ and the expectation of the right hand side is $E\{f(\omega)|\omega_0 = m''\}$, so the conditional expectations are decreasing in m. If $g(m)$ is not constant, there is a state m^* such that $c(m) < c(n)$ for all $m \leq m^* < n$. For any (x_t, y_t) path such that for some interval of time, $x_t \leq m^* < y_t$, the inequality is strict. The set of such paths has positive probability, and so the inequality between conditional expectations is strict. □

The first application of this Lemma 2 compares the present discounted value of the flow cost of being tagged from one criminal opportunity to the next. An *event* is the arrival of either the next criminal opportunity or thenext untagging. The time to the next event, τ, is distributed exponentially with parameter $p+g$. Define

$$\begin{aligned} C(m) &= E_{\tau,m}\left\{\int_0^{\tau} e^{-rt}c(m_t)dt \,\middle|\, m_0 = m\right\} \\ &= E_m\left\{\int_0^{\infty} e^{-(r+p+g)t}c(m_t)dt \,\middle|\, m_0 = m\right\} \end{aligned}$$

To save space below, any expectation operator containing m in the subscript will denote an expectation conditional on the event $\{m_0 = m\}$. Everything to the right of the vertical line will be surpressed.

Lemma 3. *$C(m)$ is non-decreasing in m, and strictly increasing in m if $c(m)$ is not constant.*

Proof. Integrating by parts, $C(m) = \int_0^{\infty} e^{-(p+g+r)t}c(m_t)dt$. The conclusion follows from Lemma 2. □

The next application compares the present discounted value of non-flow costs realized at events.

Lemma 4. *If $\{m_t\}_{t=0}^{\infty}$ is a birth-death process, $f(m)$ is a non-increasing function of m for each t and τ is the arrival time of the next event, then the conditional expectation $E_{m,\tau}\{e^{-r\tau}f(m_\tau)\}$ is non-increasing in m.*

Proof. This is another application of Lemma 2.

$$E_{m,\tau}\{e^{-r\tau}f(m_\tau)\} = (p+g)\int_0^{\infty} e^{-(p+g+r)t}f(m_t)dt$$

and the conclusion follows from the Lemma. □

Suppose an individual has a decision opportunity at time 0. Let τ denote the time to the next event and σ denote the time to the next criminal opportunity after τ. Then τ and σ are distributed independently and exponentially, τ with parageter $p+g$ and σ with parameter p. Let $V(m,\delta)$ denote the value of optimal choice for an untagged individual in state m with net expected reward δ, and let $W(m,\delta)$ denote the same reward for a tagged individual. For any function $f(m,\delta)$, let $\hat{f}(m) = E_\delta f(m,\delta)$, the result of expecting out u (which is independent of all other random variables in the model). Define $V_x(m,\delta)$ to be the value to an untagged individual of making decision $x \in \{C, N\}$ (Commit or Not) and continuing optimally. Define $W_x(m,\delta)$ similarly for tagged individuals.

Begin with tagged individuals:

$$
\begin{aligned}
W_C(m,\delta) &= \delta - C(m) + \\
&\qquad E_{\tau,m}\Big\{e^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\} + \frac{p}{p+g}\hat{W}(m_\tau)\Big)\Big\} \\
W_N(m,\delta) &= 0 - C(m) + \cdots
\end{aligned}
$$

If the individual commits the crime, she receives expected immediate net reward δ. She also pays a flow cost of being tagged until the next event. The expected value of this cost is $C(m)$. With probability $g/(p+g)$ that event is an untagging. She then waits, without paying tagging costs, until the next crime opportunity, at which time she plays optimally. With probability $g/(p+g)$ that event is a criminal opportunity, and she plays optimally, still tagged. This gives $W_C(m,\delta)$. A similar explanation covers $W_N(m,\delta)$.

The one-step deviation principle implies that the optimal strategy for a tagged player has

$$
\sigma_T(m,\delta) \in \begin{cases} \{C\} & \text{if } \delta > 0, \\ \{C, N\} & \text{if } \delta = 0, \\ \{N\} & \text{if } \delta < 0. \end{cases}
$$

Furthermore

$$
\begin{aligned}
W(m,\delta) &= \max\{W_C(m,\delta), W_N(m,\delta)\} \\
&= \max\{\delta, 0\} - C(m) + \\
&\qquad E_{\tau,m}\Big\{e^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\} + \frac{p}{p+g}\hat{W}(m_\tau)\Big)\Big\}
\end{aligned}
$$

Now consider untagged individuals:

$$\begin{aligned} V_C(m,\delta) = {} & \delta - qC(m) + \\ & E_{\tau,m}\Big\{qe^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\big\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\big\} + \frac{p}{p+g}\hat{W}(m_\tau)\Big) + \\ & (1-q)e^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\big\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\big\} + \frac{p}{p+g}\hat{V}(m_\tau)\Big)\Big\} \qquad (6) \end{aligned}$$

$$V_N(m,\delta) = E_{m,\tau}\Big\{e^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\big\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\big\} + \frac{p}{p+g}\hat{V}(m_\tau)\Big)\Big\} \qquad (7)$$

In the first instance the individual commits the crime and gets expected immediate net return δ. With probability q she is tagged, and her life continues on as a tagged individual, just as in $W_x(m,\delta)$. With probability $1-q$ she is not tagged and merely waits for the next criminal opportunity, at which she plays optimally. That piece is broken down into its τ and σ components in order to facilitate comparisons with the other value functions. A similar argument gives $V_N(m,\delta)$. Finally, $V(m,\delta) = \max\{V_C(m,\delta), V_N(m,\delta)\}$.

Define $\Delta^T(m,\delta) = W(m,\delta) - V(m,\delta)$. A calculation shows that:

$$\begin{aligned} V(m,\delta) = {} & \max\Big\{\delta - qC(m) + \frac{pq}{p+g}E_{\tau,m}e^{-r\tau}\hat{\Delta}^T(m_\tau), 0\Big\} + \\ & E_{m,\tau}\Big\{e^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\big\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\big\} + \frac{p}{p+g}\hat{V}(m_\tau)\Big)\Big\} \\ W(m,\delta) = {} & \max\{\delta,0\} - C(m) + \\ & E_{m,\tau}\Big\{e^{-r\tau}\Big(\frac{g}{p+g}E_\sigma\big\{e^{-r\sigma}\hat{V}(m_{\tau+\sigma})\big\} + \frac{p}{p+g}\hat{W}(m_\tau)\Big)\Big\} \\ \Delta^T(m,\delta) = {} & \max\{\delta,0\} - C(m) - \max\Big\{\delta - qC(m) + \frac{pq}{p+g}E_{\tau,m}e^{-r\tau}\hat{\Delta}^T(m_\tau), \\ & 0\Big\} + \frac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{\Delta}^T(m_\tau) \\ = {} & \max\{\delta,0\} + \Big(\frac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{\Delta}^T(m_\tau) - C(m)\Big) - \\ & \max\Big\{\delta + q\Big(\frac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{\Delta}^T(m_\tau) - C(m)\Big), 0\Big\} \qquad (8) \end{aligned}$$

The optimal policy for an untagged player has:

$$\sigma_U(m,\delta) \in \begin{cases} \{C\} & \delta + q\Big(\dfrac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{\Delta}^T(m_\tau) - C(m)\Big) > 0\,, \\ \{C,N\} & = 0, \\ \{N\} & < 0\,. \end{cases} \tag{9}$$

The strategy σ_T is clearly a monotonic reservation strategy. The strategy σ_U will be monotone if and only if the left hand side of the condition is non-increasing in m, and a reservation strategy if and only if it is increasing in δ. Lemma 2 states that $C(m)$ is increasing in m, so the monotonicity of σ_U will follow from the independence of $\{m_t\}$ and τ if $\Delta^T(m,\delta)$ is non-increasing in m for all δ. The reservation property is an immediate consequence of $\Delta^T(m,\delta)$'s being non-decreasing in δ.

Let S denote the set of all functions $f(m,\delta)$ which are non-positive, non-increasing in m and non-decreasing in δ. Define the map T on S such that

$$\begin{aligned} Tf(m,\delta) = \max\{\delta,0\} + \Big(\frac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{f}(m_\tau) - C(m)\Big) - \\ \max\Big\{\delta + q\Big(\frac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{f}(m_\tau) - C(m)\Big), 0\Big\} \end{aligned} \tag{10}$$

The difference $\Delta^T(m,\delta)$ is a fixed point of the map T. Properties of Δ^T are inherited from the map T.

Lemma 5. *The map T maps S onto S is increasing and is a contraction. T is non-decreasing in r and F, and non-increasing in v and c.*

Proof. Obviously T is increasing in f for any q less than 1. Suppose that $f \in S$. Let

$$\phi = \frac{p}{p+g}E_{\tau,m}e^{-r\tau}\hat{f}(m_\tau) - C(m)\,.$$

First observe that ϕ is non-increasing in m. Depending upon the second max, the value of $Tf(m,\delta)$ is either $\max\{\delta,0\}+\phi$ or $\max\{\delta,0\}+(1-q)\phi$. In either case, Tf must be non-decreasing in m and increasing in δ, so $T : S \to S$.

To see that T is a contraction map, consider f and g in S and pick a state m and reward u. Let $z = p/(p+g)$. First, suppose that for both f and g, the maxima are the left hand terms. Then

$$Tf(m,\delta) - Tg(m,\delta) = \frac{p(1-q)}{p+g} E_{\tau,m} e^{-r\tau}\Big(\hat{f}(m_\tau) - \hat{g}(m_\tau)\Big)$$

Thus $||f-g|| < \epsilon$ implies $|Tf(m,\delta) - Tg(m,\delta)| \leq z(1-q)\epsilon$. Next, suppose that both maxes are the right hand terms, $C(m)$. Then $||f-g|| < \epsilon$ implies $|Tf(m,\delta) - Tg(m,\delta)| \leq z\epsilon$. Next, suppose that the max for f is the left hand term and the max for g is the right hand term. That is,

$$\delta - qC(m) + \frac{pq}{p+g} E_{\tau,m} e^{-r\tau} \hat{f}(m_\tau) \geq 0$$
$$\geq \delta - qC(m) + \frac{pq}{p+g} E_{\tau,m} e^{-r\tau} \hat{g}(m_\tau)$$

Calculating with this inequality shows that if $||f-g|| < \epsilon$, then

$$(1-q)z\epsilon \leq ||Tf - Tg|| \leq z\epsilon$$

The same result obtains if the max for f is on the right and for g on the left. Thus the operator T on S contracts at rate $z < 1$.

The remaining effects of parameter changes are straightforward calculations. The effect of increasing F follows from the fact that if $h(x)$ is non-decreasing in x, then the expectation Eh is non-decreasing as the probability distribution increases in the sense of first-order stochastic dominance. □

All the existence and comparative statics results will be derived from the Tarski's fixed point theorem for increasing functions on complete lattices. The required monotonicity is a consequence of the following lemma:

Lemma 6. *For all f in S, Tf is non-decreasing in the death rates of the $\{m_t\}_{t\geq 0}$ process.*

The proof is another kind of coupling argument.

Proof. It suffices to show that ϕ is non-decreasing in the death rates of m. Since the random time τ is independent of the m_t process, it suffices to show that for any time T and and non-increasing function $f(m)$, the expectation $E_m f(m_T)$ is non-decreasing in the death rates μ_m.

Let $\{\lambda^x_m, mu^x_m\}_{m=0}^{N-1}$ and $\{\lambda^y_m, mu^y_m\}_{m=0}^{N-1}$ be two sets of birth and death rates such that $\lambda^x_m \geq \lambda^y_m$ and $\mu^x_m \leq \mu^y_m$. Thus the first set has higher birth and lower death rates than the second. Let $\lambda_m = \max\{\lambda^x_m, \lambda^y_m\}$ and $\mu_m = \max\{\mu^x_m, \mu^y_m\}$. (The Lemma only requires $\lambda^x_m = \lambda^y_m$, but I may find this fact useful later.)

Construct the coupling $\{x_t, y_t\}_{t\geq 0}$ as follows. The processes begin in the same state, that is, $x_0 = y_0$. Whenever $x_t = y_t = m$, births arrive at rate λ_m and deaths arrive at rate μ_m. When a birth arrives, process x increments with probability λ^x_m/λ_m and y increments with probability λ^y_m/λ_m. Whenever $x_t \neq y_t$, the two processes evolve independently according to their respective birth and death rates. It is easy to see that each marginal process $\{x_t\}_{t\geq 0}$ and $\{y_t\}_{t\geq 0}$ evolves according to its own birth and death rates, and that almost surely, $x_t \geq y_t$. Consequently, almost surely $f(x_t) \leq f(y_t)$, and so this is true in expectation as well. □

This is enough to prove theorems 1 and 2. The existence of equilibrium will follow from Zhou's (1994) extension of Tarski's fixed point theorem. We will follow Topkis (1998). It is also convenient to use lattice arguments to get the effects of parameter changes on Δ^T even though the existence of a fixed point for T is guaranteed by the Contraction Mapping Theorem.

First, the set S is a complete lattice under the pointwise "at least as big as" order. It follows from the fixed point theorem and Lemma 5 that T has fixed points, the set of fixed points can be totally ordered, and that the least and greatest fixed points are non-decreasing in r and F, and non-decreasing in v and c. It follows from Lemma 6 that the greatest and least fixed points are non-decreasing in the death rates of the m_t process. Since T is a contraction, it has in fact only one fixed point, Δ^T, and Δ^T varies as just described with the parameters.

Since $\Delta^T(m, \delta)$ is non-increasing in m and since the time τ is independent of the m_t process, it follows from lemma monotone that ϕ is non-increasing in m_0. If $c(m)$ is not constant it is strictly increasing in m. From equation 9 it follows that the optimal policies for an untagged player are all monotonic reservation strategies.

Although we cannot totally order all strategies, the previous argument shows that all best responses are monotonic strategies, that σ_U miyes in at most one state, and that $\sigma_T \equiv 1$. The set of all such strategies is totally ordered by $\succeq$, the "more criminal than" ordering. Each such σ can be characterized by a pair (m, p) for σ_U with $p > 0$ such that $\sigma_U(m')$ is 0 for $m' > m$, 1 for $m' < m$ and p for $m' = m$. Then $\sigma \succeq \sigma'$ if and only if $m' > m$ or $m' = m$ and $p' \geq p$.

Now that the strategy set has been reduced to degrees of criminality, we can see how "being more criminal" provides a strategic complementarity. We will demonstrate that individual i's best response correspondence $B(\sigma_U)$ is increasing in the order $\succeq$. Define

$$B(\sigma_U) = \begin{cases} \{1\} & \text{if} \quad \delta + q\Big(\dfrac{p}{p+g} E_{\tau,m} e^{-r\tau} \hat{\Delta}^T(m_\tau) - C(m)\Big) > 0\,, \\ [0,1] & \text{if} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = 0, \\ \{0\} & \text{if} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad < 0\,. \end{cases}$$

This correspondence is increasing in σ. If $\sigma' \succeq \sigma$, then as equations (1) and (2) show, the birth rates for the state process remain the same and the death rates increase. It follows from Lemma 6 that ϕ increases, and therefore so does the threshold state. Similarly $B(\sigma)$ is increasing in r and F, and decreasing in v, and c.

Zhou's fixed point theorem requires that $B(\sigma_U)$ is increasing in its argument, which we have shown, and that $B(\sigma_U)$ has the lattice property of being sub-complete. This is guaranteed by the fact that the strategy set is compact in the natural product topology and totally ordered, and that the ordering is continuous in the natural topology.[19] Consequently the best-response correspondence has a fixed point, and any such fixed point is an equilibrium. This proves Theorem 1. The comparative statics result, Theorem 2, follows immediately from Topkis (1998, Theorem 2.5.2) and the comparative statics of ϕ.

Proof of Theorem 3 and the Corollary: The birth rates for the population process are fixed by the parameters of the model. Only the death rates depend on the strategy, and they are non-decreasing in the criminality of the strategy. We are given odds ratios r_n for each adjacent pair of $N+1$ states ordered from 0 to N: $\pi(n) = \pi(n^-)r_n$, and we want to infer that if one r_i increases, then for any number A the probability $\Pr\{\tilde{n} \geq A\}$ does not decrease. To see this, write

$$\Pr\{\tilde{n} \geq A\} = \frac{\sum_{l=A}^{N} r_1 \cdots r_l}{1 + \sum_{k=1}^{N} r_1 \cdots r_k}$$

Differentiating with respect to r_i shows that $\Pr\{\tilde{n} \geq A\}$ increases in r_i. Making a strategy more criminal does not lower and can raise a death rate, which (weakly) decreases some odds ratios. The Theorem follows from Theorem 2.

To prove the corollary, observe that the long run crime rate $\eta(n)$ decreases in n. The expectation of a decreasing function increases as the distribution falls with respect to first order stochastic dominance. □

6.2 Theorem 4

The expected present value of the cost of being caught is bounded above by $qv+qc(0)/r$. Any crime opportunity with a reward u exceeding this value will be accepted. The hypothesis of the theorem is that the occurrence of such opportunities is a positive probability event. Thus in each state, the death rate is at least $pq\big(1 - F(qv + qc(1)/r)\big) > 0$. On the other hand, as $g \to 0$ the birth rates are converging to 0 in each state. In the limit process, 0 is the only recurrent state, and the unique ergodic distribution puts all its mass at 0. In this limit, every individual is tagged, and so equilibrium behavior is given by σ^*. For g sufficiently small, the fraction of criminal opportunities coming to untagged individuals is very small. Since the birth rates are very small, the invariant distribution, given by (3), is nearly point mass at 0. The long run crime rate is continuous in the parameter values and with respect to the invariant distribution. At $g = 0$, it is η^*.

As g becomes large, $C(m)$ converges to 0, and so it is apparent from equations (6) and (7) that $V_C(m,\delta) - V_N(m,\delta)$ converges to δ for all m. Thus for each state the threshold utility converges to qv. Since qv is a continuity point of F, the death rate in each state converges to $pq\big(1 - F(qv)\big)$, the death rate which would result from decision rule σ^*. The birth rate in each state is converging to 0, and so in the limit, the invariant distribution has point mass at 1, the equilibrium strategy has $\sigma_C = \sigma^*$, and the long run crime rate is η^*.

6.3 Computing Equilibria

Best responses to a given strategy $\rho = (\rho_U, \rho_T)$ used by the population are determined by the sign of the expression

$$\delta + q\Big(\frac{p}{p+g} E_{\tau,m} e^{-r\tau} \hat{\Delta}^T(m_\tau) - C(m)\Big)$$

The dependence of this expression on ρ is through the evolution of the state process m_t, which appears both in the term $E_{\tau,m} e^{-r\tau} \hat{\Delta}^T(m_\tau)$ and in the term $C(m)$. The operator T defined in equation (10), whose fixed point is Δ^T is a contraction map, so in principle one should be able to compute the function for various values of ρ. The trick to the computation is to get an expression for the operator. This requires computing, for a given function $f(m, \delta)$, the expression $E_{\tau,m} \hat{f}(m_\tau)$, and computing for the cost function c the expression $E_{\tau,m} \int_0^\tau e^{-rt} c(m_t)\, dt$. Fortunately the apparatus of birth-death processes provides a convenient computational technique for producing these expressions. To illustrate the technique, consider $C(m)$.

Begin in state m at time 0, and define a time σ which is exponentially distributed with parameter $\lambda(m) + \mu(m) + p + g$, where $\lambda(m)$ and $\mu(m)$ are the birth and death rates, respectively, for the m_t process with strategy ρ, as given by equations (1) and (2). Then

$$\begin{aligned}
C(m) &= E_{\tau,m} \int_0^\tau e^{-rt} c(m_t)\, dt \\
&= E_\sigma \int_0^\sigma e^{-rt} c(m)\, dt + E_\sigma \Big\{ e^{-r\sigma} \Big(\frac{\lambda(m)}{\lambda(m) + \mu(m) + p + g} C(m^+) + \\
&\qquad \frac{\mu(m)}{\lambda(m) + \mu(m) + p + g} C(m^-) \Big) \Big\} \\
&= \frac{1}{\lambda(m) + \mu(m) + p + g} c(m) +
\end{aligned}$$

$$\frac{\lambda(m)}{\lambda(m)+\mu(m)+p+g+r}C(m^+) + \frac{\mu(m)}{\lambda(m)+\mu(m)+p+g+r}C(m^-)$$

This recursion formula says, compound the flow cost from now until time σ. At time σ either a criminal opportunity or untagging has arrived, $\sigma = \tau$, or the m_t process has moved: up with proability $\lambda(m)/\lambda(m) + \mu(m) + p + g$ and down with probability $\mu(m)/\lambda(m) + \mu(m) + p + g$. In the first case, stop compounding. In the second and third cases, just add on $C(m^+)$ or $C(m^-)$, appropriately discounted.

This formula for $C(m)$ suggests examining the operator $O(f)$ defined such that:

$$O(f)(m) = \frac{1}{\lambda(m)+\mu(m)+p+g}c(m) + \frac{\lambda(m)}{\lambda(m)+\mu(m)+p+g+r}f(m^+) + \frac{\mu(m)}{\lambda(m)+\mu(m)+p+g+r}f(m^-)$$

This operator is a contraction on the space of functions from Ω to $\mathbf{R}$, and its fixed point is $C(m)$. This operator can easily be iterated on the computer to approximate $C(m)$. A similar operator can be iterated to find $E_{\tau,m}\hat{f}(m)$ for a given function f. With these functions in hand, $T(f)$ can be computed. And iterating T gives Δ^T, from which best responses and equilibria are easily computed.

Notes

[1]Cowell (1990), McGraw and Scholz (1992) and Schwartz and Orleans (1966) are a few of the many studies demonstrating this effect.

[2]Given decision rules for all individuals, the process $\{n_t\}_{t=0}^{\infty}$ describing the number of untagged individuals in the entire population can be constructed similarly.

[3]It makes no qualitative difference whether we assume this or $\mu = \mu^T$.

[4]Other parameter values are: $p = 0.7$, $g = 0.002$ $c(m) = 2.5m$, $r = 0.15$, $\epsilon = 0.01$, $u_h = 5.0$ and $u_l = 1.0$. In this plot, $N = 200$.

[5]Other parameter values are: $p = 0.65$, $q = 0.95$ $c(m) = 2.0m^4$, $r = 0.01$, $\epsilon = 0.01$, $u_h = 5.0$ and $u_l = 1.0$. In this plot, $N = 200$.

[6]See Book (1999) and Brilliant (1989).

[7]See Blume and Durlauf (1998).

[8]This is closely related to statistical discrimination models.

[9]For instance, if sufficiently few individuals are marked most of the time, the incidence of the mark may fall below the threshold of social attention which would invoke a stigmatic response.

[10]Lessig (1995, p. 965).

[11]See Book (1999) and Brilliant (1989) for startling examples of this recent phenomenon.

[12]21 U.S.C. §862(a). Those wishing to remove the ban must be able to "prove" rehabilitation, a difficult and potentially costly undertaking. Individual states may opt out of this provision. As of January 2002, 19 states have left the ban in place, and some of those states which have modified the ban have in fact extended its coverage. See Hirsch (2002, Chapter 1).

[13]15 U.S.C. 1681c(a)(5). See Hirsch (2002, Chapter 2).

[14]See Goffman (1963, p. 12.).

[15]Newman (1999, p. 104). "Burger Barn" is a pseudonym for a national chain of fast food restaurants.

[16]For instance, the Anti-Drug Abuse Act of 1988 punished simple possession of five grams of crack with a mandatory-minimum sentence of sixty months in prison. Possession of any quantity of any other substance by a first-time offender is a misdemeanor punished by a maximum of twelve months in prison. See 21 U.S.C. §844(a) (1994). One estimate has it that

nearly one-third of young male African-Americans between the ages of 20 and 29 are in prison, jail, or on probation or parole. See Mauer and Huling (1995).

[17] Beginning in 1993 the New York City Police Department began aggressive enforcement of "public order" violations such as public drunkenness, prostitution, aggressive panhandling and the like. City officials and some criminologists believe this strategy is responsible for New York's above average rate reductions burglaries, robberies and murder. See Kahan (1997).

[18]The exceptional set is Matsui and Matsuyama (1995), Hofbauer and Sorger (2000), and for stochastic population games, Blume (1995).

[19]It has to be "subcomplete". See Topkis (1998), Theorem 2.5.1. The strategy space itself must also be a complete lattice, which is guaranteed again by order-continuity, its compactness, and that the order is complete.

References

Anderson, E. (1990): *Streetwise: Race, Class, and Change in an Urban Community.* University of Chicago Press, Chicago.

Becker, G. S. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169–217.

Blume, L. E. (1993): "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 4, 387–424.

Blume, L. E. (1995): "Evolutionary Equilibrium with Forward-Looking Players," Cornell University.

Blume, L. E., and S. Durlauf (1998): "Equilibrium Concepts for Social Interaction Models," forthcoming, *International Game Theory Review.*

Book, A. S. (1999): "Shame on You: An Analysis of Modern Shame Punishment as an Alternative to Incarceration," *William and Mary Law Review*, 40, 653–686.

Brilliant, J. A. (1989): "The Modern Day *Scarlet Letter:* A Critical Analysis of Modern Probation Conditions," *Duke Law Journal*, pp. 1357–1384.

Cowell, F. A. (1990): *Cheating the Government: The Economics of Evasion.* MIT Press, Cambridge MA.

Ethier, S. N., and T. G. Kurtz (1986): *Markov Processes: Characterization and Convergence.* Wiley-Interscience, New York.

Goffman, E. (1963): *Stigma: Notes on the Management of Spoiled Identity.* Simon and Schuster, New York.

Hirsch, Amy E., et. al. (2002): "Closing the Door: Barriers Facing Parents with Criminal Records," Center for Law and Social Policy, available at http://www.clasp.org.

Hofbauer, J., and G. Sorger (2000): "A Differential Game Approach to Equilibrium Selection," *International Game Theory Review*, forthcoming.

Kahan, D. M. (1997): "Social Influence, Social Meaning, and Deterrence," *Virginia Law Review*, 83, 349–95.

Kandori, M., G. Mailath, and R. Rob (1993): "Learning, Mutation and Long Run Equilibrium in Games," *Econometrica*, 61, 29–56.

Kingman, J. F. C. (1993): *Poisson Processes.* Oxford University Press, New York.

Lessig, L. (1995): "The Regulation of Social Meaning," *University of Chicago Law Review*, 62, 943–1045.

Link, B. G., and J. C. Phelan (2001): "Conceptualizing Stigma," *Annual Review of Sociology*, 27, 363–85.

Matsui, A., and K. Matsuyama (1995): "An Approach to Equilibrium Selection," *Journal of Economic Theory*, 65, 415–434.

Mauer, M., and T. Huling (1995): *Young Black Americans and the Criminal Justice System: Five Years Later*. The Sentencing Project, New York, NY, Report #9070.

McGraw, K. M., and J. T. Scholz (1992): "Taxpayer Adaptation to the 1986 Tax Reform Act: Do New Tax Laws Affect the Way Taxpayers Think About Taxes?," in *Why People Pay Taxes: Tax Compliance and Enforcement*, ed. by J. Slemrod. University of Michigan Press, Ann Arbor.

Newman, K. S. (1999): *No Shame in My Game: The Working Poor in the Inner City*. Random House, New York.

Sampson, R. J., and W. B. Groves (1989): "Community Structure and Crime: Testing Social Disorganization Theory," *American Journal of Sociology*, 94(4), 774–802.

Schwartz, R. D., and S. Orleans (1966): "On Legal Sanctions," *University of Chicago Law Review*, 34, 296–99.

Topkis, D. M. (1998): *Supermodularity and Complementarity*. Princeton University Press, Princeton, NJ.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61, 57–84.

Zhou, L. (1994): "The set of Nash equilibria of a supermodular game is a complete lattice," *Games and Economic Behavior*, 7, 295–300.

Author: Lawrence Blume

Title: Stigma and Social Control

Reihe Ökonomie / Economics Series 119

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

ISSN: 1605-7996