Professor Yoram Margalioth
Tel Aviv University

WELFARE STIGMA

Wednesday, February 13, 2008
12:30 – 2:00
Solarium (Room FA2)
84 Queen’s Park

For workshop purposes only.
Please do not cite or distribute without the authors’ permission.
The Role of Stigma in the
Design of Welfare Programs

Tomer Blumkin*  Yoram Margalioth**  Efraim Sadka***

February 04, 2008

Abstract
We consider the notion of welfare stigma à la Besley and Coate (1992b). This stigma is attributed by society to welfare claimants in case they are perceived undeserving in the sense that they falsely claim to be eligible for welfare benefits. However, due to imperfect information, this stigma may be extended, with some probability, to all welfare claimants. We examine the implications of this kind of stigma for the design of welfare programs.

JEL Classification: H2, D6

Key Words: Welfare, Take Up, Targeting, Inequality, Stigma.
1. Introduction

The goal of most welfare programs is to provide benefits to the least well-off individuals. These individuals are usually characterized by low earning ability or ill health – information that cannot be directly observed by the government, nor can be easily verified.\(^1\) To overcome these problems the government uses various screening devices. Direct devices include means-testing by reviewing documentation, conducting interviews, and testing by specialists. Indirect screening include: targeting groups \[^{[tagging \text{ a la Akerlof (1978)}]}\]: basing welfare eligibility on observable characteristics such as old age, education or observable disability, correlated with ability; targeting benefits: offering in-kind benefits (e.g., wheelchairs) that deserving individuals, namely, the intended beneficiaries of the program, would find relatively more attractive \[^{[Nichols and Zeckhauser (1982)]}\]; and welfare ordeals: adding requirements, such as work or training requirements to the program (even if they entail pure deadweight costs) that undeserving individuals would find relatively costly, hence self-select out of the program \[^{[Nichols and Zeckhauser (1982); Besley and Coate (1992a), (1995)]}\].

In designing the above-mentioned screening devices the government faces a fundamental trade-off between benefits take-up, that is, securing that the eligible individuals will actually receive the benefits;\(^2\) and targeting efficiency, namely, excluding non-eligible claimants.\(^3\)

\(^{1}\) Mental problems and back pains are the two leading causes for inability to work stated in claims for government income support. Unfortunately these causes are especially difficult to verify \[^{[King (2004); see also Campolieti (2002)]}\].

\(^{2}\) Ample empirical research has shown that complex and time-consuming application processes \[^{[e.g., Moffitt (2003); Currie (2006)]}\], as well as stigma \[^{[Moffitt (1983)]}\] significantly reduce welfare take-up rates. Lack of information about the existence of some programs and their terms or eligibility rules is another often-mentioned explanation for low take-up rates.

\(^{3}\) A well-known example is the ‘Dutch disease’ of the 90s. In 1990, the proportion of Dutch workers aged 15 to 44 living on disability insurance was about three times higher than in other European countries. In that year, Prime Minister Lubbers called the Netherlands a "sick" country, and vowed to
Stigma is a major cause for low take up [see, inert-alia, Moffitt (1983), Hancock et al. (2004) and Pudney et al. (2006)] and reducing welfare stigma is a commonly stated policy objective (e.g., Besley and Coate (1992b), (1995)]. A recent example of a policy change in this direction is the introduction of Electronic Benefit Transfer (EBT) systems to provide food stamps program participants with a magnetic debit card that looks more like a regular debit or credit card.

In this paper we draw a distinction between two different explanations for the phenomenon of welfare stigma. Stigma is a Greek word that referred to a kind of tattoo mark that was cut or burned into the skin of criminals, slaves, or traitors in order to visibly identify them as blemished or morally polluted persons [Frey (2003)]. Sociological theory defines stigma as a phenomenon whereby an individual with an attribute, which is deeply discredited by his/her society, is rejected as a result of that attribute [Goffman (1963)]. The first, well-known, explanation for welfare stigma is that welfare recipients are publicly exposed as being unable to support themselves by their own resources. The relevant attribute is therefore inability, which is an accurate description of the targeted welfare recipient from the government’s perspective. If this indeed is the cause of stigma it means that an individual who is unable to provide for him/herself is reduced in our minds from a whole and usual person to a tainted, and discounted one. Explained in this way, welfare stigma is an unfortunate cost that must be minimized by making the process of application and collection of welfare benefits as secretive as possible or by providing the benefits universally.

A second possible explanation for welfare stigma is that the relevant attribute is not genuine inability but rather laziness or dishonesty. According to this explanation, society in general is sympathetic towards the unfortunate disabled and “cure” it. The Netherlands was indeed “cured” in 1993 when greater scrutiny was introduced [Aarts and De Jong (1996)].
wishes to help them through welfare programs and otherwise (Wolf, 2004). However,
in practice, there may exist also people who pretend to be deserving, and actually
manage to receive unwarranted welfare benefits. Because of this phenomenon,
welfare recipients are viewed with some suspicion. According to this explanation, the
cause of welfare stigma is that the welfare recipients “are held responsible for their
condition” (Pettigrew, 1980). This is especially true in the US. Most Americans,
elites, middle class and even the poor, believe that effort is what principally accounts
for how people do in life, and that those who are poor simply have not tried hard
enough (Rainwater, 1974). A systematic investigation of American conceptions of the
poor (Feagin, 1972) demonstrates that persons are believed to be poor primarily as a
result of their own failings. Studies comparing US and European style welfare states
suggest that one of the major reasons that Americans redistribute less than Europeans
is that Americans believe that they live in an open and fair society and that if someone
is poor it is their own fault [Alesina, Glaesar and Sacerdote, (2001); Alesina and
Angeletos (2005)]. Similar views, however, exist in most of Europe too, although to a
lesser extent (e.g., Commission of the European Communities, 1977). The United
Kingdom and Ireland are the most similar to the US; thus, British and Irish
respondents chose “laziness and lack of will power” most often as a chief cause of
poverty. Even in Sweden, beneficiaries of unemployment insurance benefits were
vulnerable to the same accusations of laziness applied to Americans in similar
circumstances [Rainwater (1982)]. Because of Sweden’s preoccupation with problems
of male alcoholism, the popular imagery of welfare chiselers applied more to men
than to women [Rainwater (1979)].

This second explanation for welfare stigma was carefully developed by Besley
and Coate (1992b) as follows. Society cares about the welfare of the poor and is
willing to transfer resources to them, but only if the claimants are "deserving." Undeserving claimants are individuals that can support themselves but nevertheless choose to claim benefits at the expense of society at large. There is no way of distinguishing between deserving and undeserving claimants. Therefore, society attributes to all welfare claimants some stigma cost. This is a form of statistical discrimination (profiling) and results in adverse selection: the existence of unobserved "lemons" inflicts a cost on each and every individual member of the group.

At the individual level, the cost of stigma is determined by the combination of two elements: (i) being known as a welfare claimant; (ii) being perceived as having the ability to support oneself. The former depends on the level of exposure, which is a policy instrument controlled by the government, e.g., asking claimants to undergo procedures such as reporting in person to an agency every once in a while, standing in line at welfare agency offices or in especially designated lines in the supermarket. The latter, namely, being perceived as being able to support oneself, depends on the intimate knowledge of the people around the claimant regarding her innate earning ability.

People care more about what others think about them if those others are sufficiently close to them. They care much less about what strangers think of them. Hence, stigma is mostly the result of being exposed as a welfare claimant to people who know the claimant, but are not close enough to her to actually share her interests as do family members or close friends. Exposure to people like neighbors, employers, repeated service providers or acquaintances inflicts the highest social stigma.

4 Like Besley and Coate (1992b), we believe that it is unlikely that negative self-imagery can arise independently of social attitudes; stigma depends on how people view the applicant. If the whole process were to be completely secretive, for example, through the internet, no stigma would arise. Hence, exposure could affect the level of stigma thus serving as a policy instrument.
Focusing on Besley and Coate (1992b) definition of welfare stigma, we re-examine the implications of stigma for the design of welfare programs. In particular we study the role of stigma as an ordeal mechanism in mitigating the trade-off between take-up and targeting efficiency.5 We argue that stigma imposes differential costs on deserving and undeserving claimants. The reason for the differential effect is that people who know the claimant are assumed to have some information regarding her earning ability. Hence, high ability claimants are more likely to be perceived as undeserving and therefore incur higher stigma costs compared with low ability individuals. This would help to sort out the needy individuals and enhance the target efficiency of the welfare system.6

The organization of the paper is as follows. In Section 2 we present the model. In Section 3 we analyze the government problem. Section 4 concludes.

2. The Model

We present a simple model with just the key ingredients necessary to illustrate our point. Consider an economy where individuals differ in their innate productive ability, denoted by \(w\). The population is equally divided between high-ability and low-ability individuals. We let \(\bar{w}\) and \(\underline{w}\) denote the productive ability (hence hourly wage rate, assuming constant returns to scale) of the high-ability type and low-ability type, respectively; where \(\bar{w} > w > 0\). We normalize the population of each ability-type to unity, with no loss in generality. For simplicity we assume that each individual

\[\text{\footnotesize{[5] The literature on the design of welfare programs, by and large, overlooks this second type of stigma. A recent exception is Jacquet (2006) who incorporates this notion of stigma into the analysis of optimal income taxation.}}\]

\[\text{\footnotesize{[6] Kleven and Kopczuk (2005) suggest employing program complexity (such as costly elaborate screening of claimants) as an ordeal mechanism to sort out the "pretenders". However, as indicated by the literature [see e.g., Lipsky (1984), Blank and Ruggles (1996), Daponte et al. (1999), Haider et al. (2003) and Orbach (2006)], this may sort out also the neediest individuals. In contrast, employing our notion of stigma does not suffer from this drawback, due to the differential effect on deserving and undeserving claimants.}}\]
supplies one unit of labor, hence \( w \) denotes also the gross labor income of an individual of type \( w \). Finally, we assume, following Mirrlees (1971), that individual's innate ability is private information and unobserved by the government.

Individuals may be eligible for welfare benefits. To be eligible, an individual may be asked to report to the welfare agency and be possibly subject to some requirements. Such procedures entail public exposure and hence give rise to stigma costs.\(^7\)

All individuals share the same preferences represented by the following utility function:

\[
(1) \quad u(c, s, \delta) = c - \delta \cdot s,
\]

where \( c \) denotes consumption, \( s \) denotes stigma cost and \( \delta \) is an indicator function which assumes the value of one if the individual claims for welfare benefits and zero otherwise.

We assume an egalitarian government whose re-distributive preferences are represented by a Rawlsian welfare function. That is, the government aims to attain some minimal level of well being for all individuals, denoted by \( \hat{u} > 0 \), at the minimal cost. We assume that \( \hat{u} = \hat{w} \), where \( w < \hat{w} < \bar{w} \). That is, the typical high-ability individual, whose innate ability exceeds the threshold \( \hat{w} \), can attain the minimal level of well being without the need for supplementary government support. On the other hand, the low-ability individual cannot attain this minimal level of well-being without government support.

Invoking the notion of social stigma used by Besley and Coate (1992b), we assume that the stigma cost, \( s \), suffered by an individual of type \( w \) who claims for welfare, takes the following form:

\(^7\) See the discussion in footnote 4 above.
(2) \[ s(\gamma, w, x) = \gamma \cdot p(w, x) \cdot z(x), \]

where \( 0 \leq \gamma \leq \tilde{\gamma} < 1 \) is a measure of public exposure in welfare programs (determined by the government), say, the fraction of the population that knows a certain individual is claiming for benefits; \( z \) measures the disutility associated with being on welfare; and \( p(w, x) \) denotes the probability that an individual of type \( w \) is perceived to be of high-ability conditional on the fact that \( x \) individuals are claiming for welfare. The type of stigma we focus on here derives from the fact that people cannot fully distinguish between deserving and undeserving individuals and hence assign some "cost of doubt" to all welfare claimants. Note that the literature [see, e.g., Moffitt (1983)] identifies another type of stigma cost, unrelated to being perceived as potentially undeserving, but rather driven by the sheer fact that welfare recipients are unable to support themselves by their own resources. To put our argument regarding the potential desirable feature of stigma in its sharpest relief, we set aside this other type of stigma cost, without discounting its importance.\(^8\)

We plausibly assume that \( p(w, x) \) is strictly increasing with respect to \( w \), strictly increasing with respect to \( x \), for \( x \geq 1 \), and satisfies \( p(w, x') = 0 \) for \( x' \leq 1 \). In words, the more productive an individual is, the more likely she will be perceived to be of higher ability, hence an undeserving welfare claimant. Thus, introducing stigma serves as a screening mechanism to sort out the needy individuals.\(^9\) Moreover, as the

---

\(^8\) The society's perception of deservedness depends on the definition of the government of what the minimal acceptable standard of living (well-being) means. The government rarely states this threshold explicitly but the latter may be inferred from means-testing criteria set by the government for welfare eligibility. We henceforth assume that the government sets the means testing criterion at \( \hat{w} \). It is assumed that the government cannot verify the level of income declared by the individuals. High-ability individuals can thus pretend to be deserving claimants, by reporting a low level of income (mimicking the low-ability individuals).

\(^9\) Note that imposing any sort of participation costs on welfare claimants may serve as self-selection screening device when marginal utility is diminishing with respect to consumption. However, we choose to focus on the enhancement of this self-selection effect through the imposition of differential costs on claimants. In order to capture the latter effect in the sharpest relief, we simplify by assuming that utility is linear with respect to consumption.
stigma cost is rising with respect to innate ability, deserving claimants will be the first to join the welfare system as they bear the lowest stigma costs. Thus, the larger the number of welfare claimants is, the higher the number of undeserving claimants turns out to be, hence the higher is the stigma cost associated with being a welfare claimant. Finally, when all individuals on welfare are unable to support themselves to attain the threshold level of well-being (that is, \( x \leq 1 \)), they are rationally perceived to be deserving claimants. Hence, stigma cost is zero for all individuals. We further assume that \( z(x) \) is strictly decreasing with respect to \( x \) and that \( z(2) = 0 \). In words, as the number of welfare claimants increases, the disutility associated with being on welfare decreases (being a welfare recipient becomes a more prevalent social norm). When all individuals are welfare claimants \( (x=2) \) there are naturally no stigma costs. Finally, we make the technical assumption that \( z''(x) < 0 \).

We naturally assume that individuals are rational in their beliefs in the sense that the probability assigned to all individuals to be an undeserving claimant is indeed given by the actual number of undeserving claimants. Formally,

\[
(3) \quad p(\bar{w}, x) + p(w, x) = x - 1.
\]

It is easy to provide a simple micro-foundation for the function \( p(w, x) \). We will stick to this micro-foundation in what follows. Suppose that, while individual's productive ability is private information, other individuals may observe some noisy signal correlated with this ability. Specifically, suppose that the signal may assume two values, high \( (H) \) and low \( (L) \), and let \( n(w) \) denote the probability that the signal realization is high when an individual is of ability \( w \). Suppose further that \( \bar{n} \equiv n(\bar{w}) > n(w) \equiv n \). Thus, the signal is informative, in the sense that a high-ability individual is more likely to obtain a high signal. Let \( \bar{q} \) and \( q \) denote the probability
of being perceived an undeserving claimant conditional on observing the signal \(H\) and \(L\), respectively. Employing Bayes’ Rule, it follows that \(\bar{q} = (x-1) \cdot q\) and \(\underline{q} = (x-1) \cdot (1-q)\), where \(q = \frac{n}{n+n} > 1/2\) measures the accuracy of the signal (as \(q\) increases the signal becomes more informative and for \(q=1\) the signal is perfectly informative). It thus follows that \(p(w,x) = n(w) \cdot (\bar{\bar{q}} - q) + q\) for \(x \geq 1\). It is easy to verify that all the properties of the model are satisfied. Substitution into (2) yields:

\[(4a)\quad s(\gamma, w, x) = \gamma \cdot (2q^2 - 2q + 1) \cdot (x-1) \cdot z(x),\]

\[(4b)\quad s(\gamma, w, x) = \gamma \cdot (-2q^2 + 2q) \cdot (x-1) \cdot z(x).\]

A welfare program is given by the pair \(<t, \gamma>\) where \(t\) denotes the transfer (a uniform benefit to which an individual is entitled) and \(\gamma\) measures the degree of public exposure (affecting the stigma entailed by claiming for benefits). A universal system is captured in our framework by the case where \(\gamma = 0\), that is no eligibility requirements are set (or alternatively requirements are discreet and do not involve public exposure), hence all individuals claim for benefits.\(^{10}\) A selective system is captured by the case where \(\gamma > 0\). In such a case, only a fraction of the individuals claims for benefits. Note, however, that even when \(\gamma > 0\), the case where all individuals claim for benefits forms an equilibrium. This derives from the fact that \(z(2)=0\), thus no stigma costs are entailed when being a welfare claimant becomes the social norm. Note further that a selective system will always attract some undeserving claimants. This derives from the fact that when only deserving claimants are on welfare, the stigma cost entailed by claiming for welfare is zero, so this can not form an equilibrium.

\(^{10}\) Even universal transfer programs, that is, non-excludible programs, do not have a 100% take up rate. However, making the programs selective, that is, making the benefits available only to a selected few as in the case of means-tested programs or programs for the disabled, significantly exacerbates the take-up problem (Currie 2006).
Obviously, an individual of type \( w \) will claim for benefits when faced with the welfare system \( <t, \gamma> \), if and only if:

\[
(5) \quad t - s(\gamma, w, x) \geq 0.
\]

Given a welfare program \( <t, \gamma> \), an equilibrium is given by the number of claimants in the program, \( 1 < x^* \leq 2 \), which satisfies:

\[
(6) \quad t \geq s(\gamma, w, x^*); \text{ with strict inequality only when } x^* = 2.
\]

3. The Government Problem

We turn next to formulate the government problem. The government is seeking to attain some minimal level of utility (for all individuals) at the minimal cost by choosing the transfer \((t)\) and the degree of exposure \((\gamma)\).\(^{11}\) We simplify by assuming that exposure is costless. Formally, the government is solving the following problem:

\[
\min_{\gamma \in [0,1], x} [t \cdot x]
\]

s.t.

\[
(i) \quad w + t - s(\gamma, w, x) \geq \hat{w},
(ii) \quad t \geq s(\gamma, w, x); \text{ with strict inequality only when } x = 2,
\]

where \( s(\gamma, w, x) \) is given by equation (4).

Constraint \((i)\) requires that the utility derived by all deserving claimants (the low-ability individuals) would exceed the minimal threshold, \( \hat{w} \). Condition \((ii)\) requires that individuals' beliefs about the deservedness of welfare claimants are self-fulfilling. It is straightforward to observe that constraint \((i)\) is binding in the optimal solution.\(^{12}\)

\(^{11}\) We compare between a selective welfare system and a universal one. We consider the case of a hybrid system which is partially selective later on.

\(^{12}\) To see this, suppose by way of contradiction that condition \((i)\) does not bind. There are two cases to consider. Consider first the case where the optimum is obtained by a universal system \((\gamma = 0)\) which
3.1 Comparing the Selective and the Universal Systems

To obtain the optimal solution for the government problem, we need to compare the optimal universal system with the optimal selective one. By the above considerations, the optimal universal system is simply given by $t = \hat{w} - w, \gamma = 0$ and $x = 2$. That is, the transfer, $t$, is set at a level, which is just sufficient to ensure that the low-ability individual would attain the threshold level of utility. The cost entailed by implementing the optimal universal system is hence, given by:

$$\text{cost}^{\text{universal}} = 2 \cdot (\hat{w} - w).$$

We turn next to the more interesting case of a selective welfare system. In such a case, $\gamma > 0$, $x < 2$ and both constraints [(i) and (ii)] are binding. It is first easy to verify, by virtue of the assumption that $z''(x) < 0$, that $\partial^2 s(\gamma, \bar{w}, x) / \partial x^2 < 0$. Thus, the expression on the right-hand side of constraint (ii) is strictly concave, hence for any degree of exposure, $\gamma$, provided that the level of transfer, $t$, is sufficiently small, there exist two values of $x$ for which constraint (ii) holds as an equality (see the figure below). Namely, there are in general two candidate equilibria for a selective welfare system. It is straightforward to verify that only one of the two candidate equilibria is locally stable (the one for which $\partial s / \partial x > 0$, given by $x_1$ in the figure). We will henceforth confine attention to the stable equilibrium.  

---

implies that $x=2$. In such a case, constraint (ii) is trivially non binding and one can slightly reduce $t$ without violating either one of the two constraints and reduce the cost of the welfare system, thus obtaining the desired contradiction. Consider next the case where the optimum is obtained by implementing a selective system (setting $\gamma > 0$), which implies that $x<2$. In such a case, condition (ii) is binding, but then, fixing $x$, the number of welfare claimants, one can slightly reduce both $\gamma$ and $t$ while maintaining constraint (ii) binding and not violating condition (i) by continuity considerations. By reducing the cost of the welfare system we obtain the desired contradiction.  

---

While the figure indicates the possibility of having no intersection, we will henceforth assume that a well defined solution of the system of two constraints [(i) and (ii)] for the two unknowns ($t$ and $x$) exists for any degree of exposure. Note that there (always) exists a third equilibrium when the degree of exposure is positive, namely an equilibrium where all individuals claim for welfare (which is also
Fixing the degree of public exposure, $\gamma$, a (locally stable) equilibrium for a selective welfare system is given by the (unique) pair $<t,x>$ that solves the following system of two equations:

\begin{align}
& t = \hat{w} - w + \gamma \cdot (-2q^2 + 2q) \cdot (x - 1) \cdot z(x), \tag{9} \\
& t = \gamma \cdot (2q^2 - 2q + 1) \cdot (x - 1) \cdot z(x). \tag{10}
\end{align}

Substituting for $t$ from (10) into (9) yields:

\begin{align}
& \hat{w} - w = \gamma \cdot (2q - 1)^2 \cdot (x - 1) \cdot z(x). \tag{11}
\end{align}

Substituting for $\gamma$ from (11) into (10) yields:

\begin{align}
& t \cdot x = (\hat{w} - w) \cdot A(q) \cdot x, \tag{12}
\end{align}

where $A(q) = (2q^2 - 2q + 1)/(2q - 1)^2$. 

locally stable). This equilibrium is essentially a universal program. To avoid trivial conclusions, we assume, however, that when a selective system is implemented the relevant equilibrium will be one where only a fraction of the individuals will claim for benefits.
The expression on the right-hand-side of (12) measures the cost entailed by implementing a selective welfare system as a function of $x$. It immediately follows from equation (12) that the cost is increasing with respect to the number of claimants. To minimize this cost, the optimal degree of exposure should be set at $\gamma$, the highest degree which is feasible.\(^{14}\) We thus conclude that the optimal selective system is given by the triplet $(\gamma, t(\gamma, \hat{w}, q))$ and $(x(\gamma, \hat{w}, q))$, where $t$ and $x$ are given by the solution to equations (9) and (10) for $\gamma = \gamma$. The cost entailed by implementing a selective system is given by:

\[
\text{cost}_{\text{selective}} = (\hat{w} - w) \cdot A(q) \cdot x(\gamma, \hat{w}, q).
\]

Comparing the cost entailed by the optimal universal system [equation (8)] and that entailed by the optimal selective system [equation (13)] implies:

\[
\text{cost}_{\text{selective}} < \text{cost}_{\text{universal}} \iff 2/A(q) > x(\gamma, \hat{w}, q).
\]

As $1 < x(\gamma, \hat{w}, q) < 2$, it is easy to verify that for sufficiently small $q$ [any $q$ for which $A(q) > 2$, noting that $A'(q) < 0$] the universal system dominates; whereas, for high enough values of $q$ [$q$ sufficiently close to 1, that is $A(q)$ sufficiently close to 1], the selective system prevails. Employing condition (11), it follows by virtue of the stability property of the equilibrium that $x(\gamma, \hat{w}, q)$ is strictly decreasing with respect to $q$. By continuity, it follows, then, that there exists a unique value of $q$, above which the selective system would dominate and below which the universal system would form the optimal program. Moreover, this threshold level of $q$ is strictly decreasing with respect to $\gamma$ and strictly increasing with respect to $\hat{w}$. The latter implies that for

\(^{14}\) To see this, note that as we focus on the stable equilibrium, the expression on the right-hand side of equation (11) is strictly increasing with respect to $x$. Thus, as we increase $\gamma$, $x$ would decrease. As, by assumption, increasing $\gamma$ is costless, one attains the optimum by setting $\gamma = \gamma$. Note that allowing for costly exposure would not change the qualitative nature of our results.
given \( \gamma \) and \( q \), the selective system would dominate for sufficiently small values of \( \hat{w} \); whereas, the universal system would dominate for large enough values of \( \hat{w} \).

The rationale for these results is straightforward. As \( q \) increases the differences in stigma across individuals with different abilities become more pronounced. This derives from the fact that individuals can more efficiently distinguish between low- and high-ability individuals, hence more- and less-deserving welfare claimants. For large enough \( q \), the gains from introducing stigma, thereby deterring high-ability individuals from being welfare claimants, outweigh the costs associated with the need to increase the transfer (to ensure participation of the low-ability individuals). As the degree of exposure (\( \gamma \)) increases this effect becomes more pronounced and the case for favoring a selective system becomes stronger. Finally, note that when the minimal level of well-being guaranteed by the government (\( \hat{w} \)) is small, the stigma entailed by the selective system is moderate and hence the targeting advantage of that system relative to a universal one renders the former socially desirable. As \( \hat{w} \) increases, the government has to raise the transfer and hence more individuals choose to join the welfare program. This, in turn, results in higher stigma costs entailed by participants, hence forcing the government to further raise the transfers. This amplification mechanism eventually renders the universal system (where transfers rise linearly with \( \hat{w} \)) socially desirable.

The result we obtain bears similarity to the result obtained by Besley and Coate (1992a). Besley and Coate justify the use of workfare as a self-selection screening mechanism to target benefits to low-ability individuals. The idea is that low-ability individuals have lower opportunity costs entailed by participating in workfare programs and hence a workfare requirement can reduce the costs of the welfare system and allow for enhanced poverty alleviation. The basic trade-off
between targeting and take-up re-appears in our formulation. On the one hand, shifting from a universal system to a selective one imposes stigma costs on participants (due to the inevitable existence of undeserving claimants). This results in a take-up problem, often discussed in the literature, and prima-facie calls for favoring a universal system (or for enacting policy measures aimed at reducing the stigma costs). On the other hand, the effect of stigma is differential. This novel aspect of our paper derives from the fact that we plausibly assume that while the government cannot observe the individual's productive ability, the latter is partially observable by the other individuals. In particular, the more productive an individual is, the more likely she will be perceived to be so. Hence, the more likely she will be tagged as an undeserving claimant. This differential effect implies that higher-ability individuals face a higher stigma costs from participating in welfare programs. Thus, introducing a selective system can attain enhanced targeting. When the differential effect is strong enough (depending on the degree of exposure feasible and the amount of information possessed by the individuals) a selective system would dominate.

3.2 The Optimality of a Hybrid System

Most welfare systems are neither purely selective nor are they completely universal. It would be interesting to see then whether a hybrid system, which is composed of a universal component and a selective program, would be socially desirable. It turns out, indeed, that for a wide range of parameters a hybrid (partially selective) system would form the optimal policy. We turn next to illustrate this point.

As demonstrated in the previous sub-section, for any minimal level of well-being set by the government, \( \hat{w} \), there exists a unique value of \( q \), for which the cost
entailed by a selective system is just equal to that associated with a universal welfare system. This value of $q$ is implicitly given by the solution to:

\begin{equation}
\frac{2}{A(q)} = x(\hat{\gamma}, \hat{w}, q).
\end{equation}

Moreover, given $q$, for any minimal level of well-being exceeding $\hat{w}$ the universal system would dominate, whereas, for any minimal level of well-being below $\hat{w}$ the selective system would prevail.

Now, fix $q$ and let $\hat{u}$ denote the target welfare threshold set by the government. Suppose that this threshold is set high enough, such that a universal system would be desirable (formally, $\hat{u} > \hat{w}$). We will show that rendering this system slightly selective would reduce the costs incurred by the government. Thus, a hybrid system would form the optimal solution.

To see this recall that the optimal universal system is given by $t^* = \hat{u} - w$, and, obviously, $\gamma = 0$. Now consider an alternative system, which is partially selective. Formally, let the universal component be denoted by $0 < t' < t^*$. Employing our derivations in the previous sub-section, the optimal selective program is given by the (unique) pair $<t, x>$ that solves the following system of two equations [similar to equations (9) and (10) above]:

\begin{align}
(16) & \quad t = \hat{u} - t' - w + \gamma \cdot (-2q^2 + 2q) \cdot (x - 1) \cdot z(x), \\
(17) & \quad t = \gamma \cdot (2q^2 - 2q + 1) \cdot (x - 1) \cdot z(x),
\end{align}

Note that condition (16) requires that the utility derived by a low-ability individual that participates in the selective system would be equal to $\hat{u} - t'$, so that combined with the universal transfer, $t'$, the minimal level of well-being, $\hat{u}$, will be attained. Following some algebraic manipulations (which replicate the derivations in the
previous sub-section and hence omitted) one obtains the following expression for the cost entailed by the optimal selective program:

\[
\text{cost}_{\text{selective}} = (\hat{u} - t' - w) \cdot A(q) \cdot x(\gamma, \hat{u} - t', q),
\]

with \( A(q) = (2q^2 - 2q + 1)/(2q - 1)^2 \).

The reduction in cost entailed by the universal component (due to the reduction in the transfer) is given by:

\[
\Delta \text{cost}_{\text{universal}} = 2 \cdot (t^* - t') = 2 \cdot (\hat{u} - t' - w).
\]

The hybrid system would dominate the universal system if-and-only-if the following condition holds:

\[
\text{cost}_{\text{selective}} < \Delta \text{cost}_{\text{universal}} \iff (\hat{u} - t' - w) \cdot A(q) \cdot x(\gamma, \hat{u} - t', q) < 2 \cdot (\hat{u} - t' - w)
\]

\[
\iff 2 / A(q) > x(\gamma, \hat{u} - t', q)
\]

The last inequality is satisfied when \( t' \) is sufficiently large [formally, when \( t' > \hat{u} - \hat{w} > 0 \), by virtue of equation (15)]. As the universal system dominates the selective one, by construction, it follows that the optimal solution must be a hybrid system.

The rationale for the result is straightforward. Even when shifting entirely from a universal system to a selective one is undesirable due to the large stigma costs entailed, replacing a small part of the universal transfer with a selective component is called for, because then the targeting advantage of the selective system would save costs without entailing much stigma.

4. Conclusions

In this paper we discuss the effect of welfare stigma on the design and magnitude of welfare programs. The type of stigma we focus on here derives from the fact that
people cannot fully distinguish between deserving and undeserving individuals and hence assign some "cost of doubt" to all welfare claimants.

We demonstrate that this type of stigma can be employed as an efficient ordeal mechanism to sort between the truly deserving welfare claimants and the undeserving "pretenders" relying on the public's partial ability (contrary to the government's lack of such ability) to distinguish between the two types of welfare claimants. This tool is likely to be efficient in circumstances where it is relatively easy for the public to distinguish between the truly deserving welfare claimants and the undeserving ones; and the minimal desired standard of living is relatively moderate. We also establish a case for a hybrid welfare system which consists of both a universal (stigma-free) component and a selective (stigma-generating) part, as most welfare systems are in practice.

The literature identifies another type of stigma cost, unrelated to being perceived as potentially undeserving, but rather driven by the sheer fact that welfare recipients are unable to support themselves by their own resources. To put our argument regarding the potential desirable feature of stigma in its sharpest relief, we set aside this other type of stigma cost, without discounting its importance. Clearly, the ultimate policy advice would depend on the relative magnitudes of the two types of stigma costs, which is likely to be culture-dependent and vary across countries; and calls for an empirical investigation.

---

15 In reality, the government may also have some limited ability to distinguish between deserving and undeserving claimants. Our assumption thus captures in an extreme manner the plausible presumption that the public has a comparative advantage (relative to the government) in distinguishing between the two types of welfare claimants.
References


