

NEW YORK UNIVERSITY SCHOOL OF LAW
COLLOQUIUM ON TAX POLICY
AND PUBLIC FINANCE
SPRING 2009

“Employing Statistical Stigma As
a Welfare Ordeal”

Yoram Margalioth
Tel Aviv University Law School
and NYU Law

February 19, 2009
NYU School of Law
Furman 120
Time: 4:00-6:00pm
Number 6

SPRING 2009
NYU TAX POLICY COLLOQUIUM

(All sessions meet Thursday 4:00-6:00 pm, Furman-120, NYU Law School)

1. January 15 – Daniel Shaviro, NYU Law School. “The Long-Term Fiscal Gap: Is the Main Problem Generational Inequity?”
2. January 22 – Alan Auerbach, Berkeley Economics Department and NYU Law School. “Understanding U.S. Corporate Tax Losses.”
<http://www.nber.org/papers/w14405.pdf>
3. January 29 – Edward Kleinbard, Joint Committee on Taxation. “A Reconsideration of Tax Expenditure Analysis.”
4. February 5 – Amy Finkelstein, MIT Economics Department, “EZ-Tax: Tax Salience and Tax Rates.”
5. February 12 – Dorothy Brown, Emory Law School. “Shades of the American Dream.”
6. **February 19 – Yoram Margalioth, Tel Aviv University Law School and NYU Law School. “Employing Statistical Stigma As a Welfare Ordeal.”**
7. February 26 – Leslie McCall, Northwestern University Sociology Department. “Americans' Social Policy Preferences in the Era of Rising Inequality.”
8. March 5 – Michael Doran, University of Virginia Law School. “Managers, Shareholders, and the Corporate Double Tax.”
9. March 12 – David Duff, University of Toronto Law School.
10. March 26 – Emmanuel Saez, Berkeley Economics Department. “Details Matter: The Impact of Presentation and Information on the Take-Up of Financial Incentives for Retirement Saving.”
11. April 2 – Lily Batchelder, NYU Law School.
12. April 9 – Mihir Desai, Harvard Business School and NYU Law School. “Investor Taxation in Open Economies.”
13. April 16 – Mitchell Kane, NYU Law School.
14. April 23 – Thomas Brennan, Northwestern Law School.

Employing Statistical Stigma as a Welfare Ordeal

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

February 10, 2009

Abstract

In this paper we draw attention to the existence of a probabilistic type of stigma according to which society attributes stigma to all welfare claimants (deserving and undeserving alike) due to the existence of imperfect information regarding individuals' productive abilities. We argue that this type of stigma could be used as an ordeal mechanism, where by imposing differential (participation) costs on welfare claimants, which are positively correlated with ability, the government can improve the lot of the truly needy. This mechanism should be traded-off against the advantages of having a universal welfare system, in general, and the costs of the traditional type of stigma, in particular, but should be taken into account when designing the optimal tax and transfer system.

JEL Classification: H2, D6, J1

Key Words: Welfare, Take Up, Informal labor market, Disability, Targeting, Inequality, Stigma.

* Department of Economics, Ben-Gurion University, Beer-Sheba 84105, Israel; CesIfO; and IZA. E-mail address: tomerblu@bgu.ac.il.

** The Buchman Faculty of Law, Tel-Aviv University, Tel-Aviv, Israel; Visiting Professor NYU School of Law. E-mail address: margalioth@exchange.law.nyu.edu.

*** The Eitan Berglas School of Economics, Tel-Aviv University, Tel-Aviv 69978, Israel; CesIfO; and IZA. E-mail address: sadka@post.tau.ac.il.

1. Introduction

A 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 it be easily verified. For example, chronic conditions such as back pain or arthritis, and mild emotional problems (e.g., ADD) are stated in claims for government support as the leading causes of inability to work [Campolieti (2002); King (2004); Weisbach (2007)]. Unfortunately, these conditions are especially difficult to verify and the problem of masquerading can be significant. A well-known example is the ‘Dutch disease’ of the 1990s. In 1990, the proportion of workers aged 15 to 44 living on disability insurance was about three times higher in the Netherlands than in other European countries. In that year, Prime Minister Lubbers called the Netherlands a "sick" country, and vowed to “cure” it. The country was indeed “cured” in 1993, when better screening was introduced [Aarts and De Jong (1996)].

The government uses various screening devices.¹ Direct devices include means-testing by reviewing documentation, conducting interviews, and testing by specialists. Indirect screening includes: *targeting groups* [*tagging*, according to Akerlof (1978)]: basing welfare eligibility on observable characteristics such as old age, level of 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, i.e., take the form of “digging holes”) that undeserving

¹ Screening is the fundamental issue in modern tax policy under the rubric of optimal taxation [Weisbach (2007)].

individuals would find relatively costly and hence, self-select out of the program [Nichols and Zeckhauser (1982); Besley and Coate (1992a) and (1995)].²

Good welfare ordeals are difficult to design. Simply imposing non-financial costs on welfare claimants could serve as a self-selection mechanism, assuming the benefits are fixed, because needy individuals will derive greater utility from the wealth transfer, as it is plausible to assume decreasing marginal utility from consumption. Moreover, it will clearly promote the social policy goal if this is defined to be income maintenance (rather than improving well-being), as in such a case, the disutility suffered by welfare claimants is not incorporated into the government objective. However, this is a cruel mechanism that is much inferior to a welfare ordeal with differential costs. If the welfare ordeal imposes greater costs on high ability individuals than on low ability individuals, the induced self-selection mechanism would be far more efficient (it would sort out the undesirable claimants without inflicting high costs on the targeted needy) and, hence, should be part of the optimal tax and transfer system.

In this paper, we argue that a certain notion of stigma, which we term statistical stigma, could serve as a welfare ordeal of the latter, warranted, type.³ Stigma is a Greek word that refers to a kind of tattoo that was cut or burned into the skin of criminals, slaves, or traitors, in order to identify them visibly as blemished or morally polluted persons [Frey (2003)]. Sociological theory defines stigma as a phenomenon whereby an individual with an attribute that is strongly discredited by his/her society is rejected because of that attribute [Goffman (1963)].

² Training programs may have a long-term desirable impact (by enhancing labor market productivity of participants; say, by improving their work and social skills) even when the short-run effect is minor or non-existent. However, these future gains are often hard to quantify and viewed by many as controversial. In any case, evaluating the gains derived by this type of program is beyond the scope of the current paper, which focuses on its screening role.

³ The traditional type of stigma, on which we elaborate below, may serve as a self-selection mechanism of the former (less desirable) type.

The first well-known notion, or explanation, for welfare stigma, which we will call traditional stigma, is that welfare recipients are publicly exposed as being unable to support themselves. 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 normal person to a tainted, discounted one. Explained in this way, welfare stigma is an unfortunate cost that should be minimized by making the process of application and collection of welfare benefits as discreet as possible or by providing the benefits universally. Much of the literature describes welfare stigma as a self-identity problem (Moffitt 1983). Individuals, especially senior citizens, tend to feel embarrassed to apply for welfare because they have internalized a social norm against living off other people and a corresponding normative pressure to earn money from work (Elster, 1989). This latter form of traditional stigma is only indirectly (if at all) affected by public exposure.

A second possible explanation for welfare stigma is that the relevant attribute is not actual inability but rather laziness or dishonesty. According to this explanation, in general, society is sympathetic toward the unfortunate disabled and wishes to help them through welfare programs and otherwise (Wolf, 2004). However, in practice, there may be people who pretend to be deserving and thereby manage to receive unwarranted welfare benefits. According to this explanation, 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 American- and European-style welfare systems 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 his own fault [Alesina, Glaesar and Sacerdote (2001); Alesina and Angeletos (2005)]. Similar views, however, exist in most of Europe as well, 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; indeed, 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)].⁴ Abundant evidence from the social sciences, surveying the US as well as European countries, has shown that generosity toward the poor is strong but conditional on the belief that the poor work hard [Williamson (1974); Hecllo (1986); Farkas and Robinson (1996); Gallop survey (1998); Gilens (1999); Miller (1999)].⁵

Due to this phenomenon, welfare recipients are viewed with some suspicion. According to Besley and Coate (1992b), because there is no way of fully distinguishing between deserving and undeserving claimants, society attributes to *all* welfare claimants some stigma. This is a form of statistical discrimination (profiling) and results in adverse selection: the existence of unobserved “lemons” (undeserving claimants) inflicts a cost on all members of the group (including the deserving claimants).

⁴ Because of Sweden’s preoccupation with problems of male alcoholism, the popular image of welfare chiselers applied more to men than to women [Rainwater (1979)].

It should be noted that even though the literature on the design of welfare programs largely overlooks this second notion of stigma, it appears to be, by far, the more significant of the two. As was shown in numerous surveys, much of what is widely believed about Americans' attitude towards welfare is wrong. Year after year, surveys show that most Americans see welfare as a desirable function of the government [see Gilens (1999) for a comprehensive analysis of the surveys]. The focus of considerable public resentment is not the principle of government support of the needy in and of itself, but rather, the perception that most people currently receiving welfare are undeserving.⁶ There is a widespread belief that most welfare recipients prefer to sit at home (or work only few hours) and collect benefits than to work hard and support themselves. If the Americans' individualistic ideology results in support for government aid for those who try hard but need help, then, clearly, more communitarian societies, such as most other Western countries, support the provision of welfare to the deserving poor. The welfare stigma found in the surveys mentioned above was entirely related to a remarkably strong belief held by the majority of the American public that most welfare recipients are undeserving.⁷ When assigned to an individual claimant, this stigma is in fact the statistical, not traditional, stigma.

⁶ For example, according to one poll cited in Gilens (1999), 74 percent of the public agrees that "criteria for welfare are not tough enough" but at the same time only 3 percent said they would oppose a one percent sales tax increase aimed at funding help to the poor. These trends are reflected in the welfare reform of 1996 and the shift from AFDC to TANF with its emphasis on the work requirement, as well as the significant increase in recent years of the EITC program that conditions welfare on labor market participation. The Food Stamp Program (FSP) has been transformed from a hated welfare program into a politically popular program largely because a greater percentage of its benefits flow to low-income wage earners. See Super (2004).

⁷ In the US, race plays a significant role in determining public support for welfare spending. Racial stereotypes of blacks as lazy remains credible for large numbers of Americans, and direct cash assistance, as provided under the Aid to Families with Dependent Children (AFDC) program that was terminated in the 1996 welfare reform, was seen as a welfare program designed to benefit blacks, who would rather receive money from the government than work (Brown, 2007).

In this paper, we choose to focus on the second (and, apparently, the empirically more relevant) type of stigma. In order to put our novel arguments in sharpest relief, we simplify our analysis by assuming that statistical stigma is the only source of stigma costs entailed by welfare claimants.⁸

We argue that, on the individual level, the cost of this statistical stigma is determined by a combination of two elements:

- (i) *Being known as a welfare claimant.* The statistical stigma depends on how people view the applicant. If the whole process were absolutely discreet; for example, through the internet, no stigma would arise. Stigma therefore depends on exposure, which is a policy instrument controlled by the government, e.g., asking claimants to undergo procedures such as periodically reporting in person to an agency, standing in line at welfare agency offices or in specially designated lines in the supermarket. The exposure does not necessarily involve significant administrative costs. It is an open empirical question whether ensuring privacy, or creating exposure, is the more costly alternative.
- (ii) *Being perceived as having the ability to support oneself.* This depends on the intimate knowledge of the people around the claimant regarding her innate earning ability. People care about what others think about them if these 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

⁸ We briefly discuss the implications of incorporating traditional stigma into our model in Section 4 below.

as do family members or close friends. Exposure to people like neighbors, employers, repeated service providers or acquaintances inflicts the highest social stigma.⁹

Focusing on Besley and Coate's (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 enhancing the targeting efficiency of the welfare system. We argue that statistical 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.¹⁰ Hence, high ability claimants are more likely to be perceived as undeserving and therefore incur higher stigma costs compared to low ability individuals. This distinction could help to sort out needy individuals and enhance the targeting efficiency of the welfare system.

We make the following two contributions to the literature. First, the literature to date mentions welfare stigma only as a cause of low take-up,¹¹ hence, reducing welfare stigma is a commonly stated policy objective [e.g., Besley and Coate (1992b)

⁹ This is true only if we assume that the people in the relevant environment adhere to a social norm that disapproves of welfare fraud. One could think that in certain relevant environments, i.e., inner cities, this would not be the prevalent norm. We may not mind, if we view living in the inner cities as a form of ordeal. It could be possible that most of the people living there are deserving poor. Even if that is not the case, stigma is not an issue at all in such environments; hence, the traditional type of stigma is not present either. This means that the policy tool we suggest may not be helpful in these cases, but it is not costly either. It could be costly if for some reason the "undeserving" are concentrated in the inner cities whereas the "deserving" are mostly present in the heterogeneous society. This is an open empirical question that should be answered in the process of designing the optimal system as part of weighing the costs of the traditional stigma against the benefits of better targeting of the system. Note that contrary to prevalent (erroneous) beliefs found in surveys [Gilens (1999)], blacks comprise only a minority (around 27 percent) of all poor people, and the majority of the poor population in the US does not reside in inner cities.

¹⁰ The kind of knowledge we assume that people have about the potential welfare claimant is negligible. We are talking about the ability to support oneself by doing simple jobs, therefore, all they need to perceive are relatively basic abilities.

¹¹ Take up refers to acceptance of government benefits by the intended beneficiaries. Ample empirical research has shown that complex and time-consuming application processes [e.g., Moffitt (2003); Currie, (2006)], as well as stigma [the seminal work is Moffitt (1983); see also, Hancock et al. (2004); Pudney et al. (2006)], 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 [Currie (2006)].

and (1995)]. A recent example of a policy change in this direction was the introduction of Electronic Benefit Transfer (EBT) systems to provide food stamp program participants with a magnetic debit card that looks more like a regular debit or credit card. Another example is the support of expanding the Earned Income Tax Credit (EITC) program, *inter alia*, as a means to eliminate welfare stigma. By drawing attention to the notion of statistical stigma, and making the plausible assumption that individuals surrounding the potential welfare claimant have better knowledge regarding her ability than the government does, we argue that welfare stigma could be employed as a welfare ordeal, using exposure as a policy tool.

Stigma is by no means the only potential welfare ordeal that can be used by the government, but as we argue below, it is a comparatively efficient one. For instance, Kleven and Kopczuk (2008) suggest in a recent paper, that program complexity (such as that associated with elaborate and costly screening of claimants) may be employed as an ordeal mechanism to sort out the "pretenders." However, as indicated in the literature [see e.g., Lipsky (1984); Blank and Ruggles (1996); Daponte et al., (1999); Haider et al. (2003) Orbach (2006), Currie 2006)], this is, unfortunately, likely to sort out the neediest individuals as well. The latter are likely to be the least capable of dealing with administrative hurdles. In contrast, employing our notion of stigma does not have this drawback, due to the differential effect on deserving and undeserving claimants. In other words, using statistical stigma imposes higher participation costs on high ability (hence less needy) claimants; whereas, employing complexity has the opposite (undesirable) sorting effect. This renders stigma (of the type we focus on, which seems the empirically relevant one) a comparatively efficient ordeal mechanism.

Our second contribution to the literature is the introduction of a new factor to consider when choosing between universal and selective transfer systems. Stigma can only be present in a selective system, hence, the last part of the paper identifies the parameters that should be measured in comparing the costs of the two systems, and reaches the conclusion that some degree of selectivity is likely to be optimal.

The organization of the rest of the paper is as follows. In Section 2, we present the model. In Section 3, we analyze the government's problem. Section 4 briefly discusses some extensions and in Section 5 we elaborate on policy implications. Finally we conclude.

2. The Model

We present a simple model with only the key ingredients necessary to illustrate our novel 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.¹² We let \bar{w} and \underline{w} denote the productive ability (hence the hourly wage rate) of the high-ability type and low-ability type, respectively; where $\bar{w} > \underline{w} \geq 0$. We normalize the population of each ability-type to unity, with no loss in generality. For simplicity, we assume that each individual supplies one unit of labor, hence w also denotes the gross labor income of an individual of type w . Individuals may be eligible for welfare benefits. To be eligible, an individual may be asked to report to the welfare agency and be subject to some requirements. Such procedures entail public exposure and hence give rise to stigma costs. To render our analysis meaningful, we assume that the government is faced

¹² There are in fact more than just two types in the economy. The other types are clearly observable as non-deserving and none of these types apply for welfare benefits. Taxing the other types would finance the benefits claimed by the two types we consider explicitly. The assumption that the two types are of equal size does not affect the qualitative nature of the results.

with a screening problem [in the spirit of Mirrlees (1971)], by letting the individual's gross labor income be private information, unobserved by the government.¹³ One convenient way to interpret our setting would be the following: Suppose that w denotes the income earned by the individual in the informal sector, which is unobserved by the government. The government would be interested in providing disability benefits to those who cannot work but is unable to distinguish between the truly (disabled) needy individuals (for whom $\underline{w} = 0$) and those who pretend to be disabled and claim for benefits showing no earnings in the formal sector (but working informally at the same time, earning $\bar{w} > 0$). The case of disability benefits lends itself to our analysis, which focuses on statistical stigma. The public in general is sympathetic toward the unfortunate disabled who are seen as deserving of government aid, but at the same time, is highly concerned about the potential for welfare frauds (individuals pretending to be disabled in order to be eligible for benefits).¹⁴

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

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

¹³ In this paper, we focus on the role of stigma and abstract from incorporating other (complementary) forms of screening, as those discussed in the introduction. This would serve to clarify the novel role of stigma in enhancing the efficiency of the welfare system.

¹⁴ Two remarks are called for. First, as mentioned in note 13 above, in this paper we choose to focus on the role of stigma as a screening device and on direct monetary transfers (or adjustments to the income tax schedule). In the context of disability, government may clearly address the screening issue by complementary means, including, among others, in-kind provision of goods (such as accommodations) and provision of public goods (such as access to public facilities). Second, in interpreting the differences in the income levels as stemming from (unobservable) disability, we focus only on one aspect of disability that calls for re-distributive policy set by the government. Able and disabled individuals also differ in their needs, reflected in the extra costs born by disabled to provide for accommodation in the work-place, commuting costs, higher health-care expenditures, etc., which affect disposable income even when wage rates are the same. Differences in needs, as those mentioned above, can be efficiently addressed by in-kind transfers (of goods valued more by the disabled) and commodity taxation (subsidizing goods that are disproportionately consumed by the disabled).

where c denotes consumption, s denotes stigma cost and δ is an indicator function that assumes the value of 1 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 minimal cost. We assume that $\hat{u} = \hat{w}$, where $\underline{w} < \hat{w} < \bar{w}$. That is, the typical high-ability individual, whose innate ability exceeds the threshold \hat{w} , can attain a minimal level of well being without the need for supplementary government support, and will henceforth be referred to as an undeserving claimant. On the other hand, the low-ability individual who cannot attain this minimal level of well being without government support will henceforth be referred to as a deserving claimant.

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 benefits, takes the following form:

$$(2) \quad s(\gamma, w, x) = \gamma \cdot p(w, x) \cdot z(x),$$

where $0 \leq \gamma \leq \bar{\gamma} < 1$ is a measure of public exposure in welfare programs (determined by the government), say, the fraction of the population that knows that a certain individual is claiming for benefits; x is the number of individuals who are claiming for welfare benefits; $z(x)$ measures the disutility associated with being an undeserving welfare claimant, 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 [Moffitt (1983) being the seminal paper] generally refers only to the traditional stigma cost, unrelated to being perceived as potentially undeserving, driven by the mere fact that welfare recipients are unable to support themselves. To put our argument regarding the potential desirable feature of stigma in sharpest relief, we set aside this other type of stigma cost, without discounting its importance.¹⁵

We plausibly assume that $p(w, x)$ strictly increases with respect to w , strictly increases with respect to x , for $x \geq 1$, and satisfies $p(w, x') = 0$ for $x' \leq 1$. In other words, the more productive an individual is, the more likely she will be perceived to be of higher ability, hence as an undeserving welfare claimant. Thus, introducing stigma serves as a screening mechanism to sort out the undeserving (pretending) individuals.¹⁶ Moreover, as the stigma cost rises 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, the higher the number of undeserving claimants, hence the higher the stigma cost associated with being a welfare claimant. Finally, when all individuals who claim for welfare benefits are truly unable to support themselves and 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 $z(x)$ strictly decreases with respect to x and that $z(2) = 0$. In other 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

¹⁵ For a brief discussion of traditional stigma see Section 4 below.

¹⁶ Note that imposing any sort of participation costs on welfare claimants may serve as a self-selection screening device when marginal utility diminishes with respect to consumption. However, we chose 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 sharpest relief, we simplify by assuming that utility is linear with respect to consumption.

($x=2$), there are naturally no stigma costs. Finally, we make the technical assumption that $z''(x) \leq 0$.¹⁷

We naturally assume that individuals are rational in their beliefs in the sense that the probability assigned to all individuals as being undeserving claimants is indeed given by the actual number of undeserving claimants. Formally (for $x \geq 1$),

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

[Note that $x - 1$ is the number of undeserving claimants (“pretenders”).]

One can provide a simple micro-foundation for the function $p(w, x)$. We will stick to this micro-foundation in what follows. Suppose that while an individual's productive ability is private information, other individuals may observe some noisy signal correlated with this ability (say, years of schooling, family background, the car she drives, etc.). Specifically, suppose that the signal may assume two values, high (H) and low (L), and let $n(w)$ denote the probability that the signal's value is high when an individual is of ability w . Suppose further that $\bar{n} \equiv n(\bar{w}) > n(\underline{w}) \equiv \underline{n}$. 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 \underline{q} denote the probability of being perceived as an undeserving claimant, conditional on observing signals H and L , respectively. Employing *Bayes' Rule*, it follows that $\bar{q} = (x-1) \cdot \bar{n} / (\underline{n} + \bar{n})$ and $\underline{q} = (x-1) \cdot (1 - \bar{n}) / (2 - \bar{n} - \underline{n})$. It thus follows that $p(w, x) = n(w) \cdot (\bar{q} - \underline{q}) + \underline{q}$ for $x \geq 1$ and zero otherwise. The accuracy of the signal is measured by the difference $\bar{n} - \underline{n} > 0$; as the difference increases, the signal becomes more informative, and when $\bar{n} - \underline{n} = 1$ (that is, $\bar{n} = 1$ and $\underline{n} = 0$), the signal is perfectly informative. To simplify the

¹⁷ A weaker assumption would suffice for our formal derivations.

exposition, without affecting the qualitative nature of our results, we let $\bar{n} = 1 - \underline{n} = q > 1/2$, where the parameter q measures the accuracy of the signal (the higher q is, the more informative the signal turns out to be). It is straightforward to verify that all the properties of the model are satisfied. Substituting the above expression for $p(w, x)$ into equation (2) yields:¹⁸

$$(4a) \quad s(\gamma, \bar{w}, x) = \gamma \cdot (2q^2 - 2q + 1) \cdot (x - 1) \cdot z(x),$$

$$(4b) \quad s(\gamma, \underline{w}, x) = \gamma \cdot (-2q^2 + 2q) \cdot (x - 1) \cdot z(x).$$

A welfare program is given by the pair $\langle t, \gamma \rangle$, where t denotes the transfer (a uniform benefit to which an individual is entitled) and γ measures the degree of public exposure (affecting the stigma entailed by claiming for benefits).¹⁹ A universal system is captured in our framework by the case where $\gamma = 0$; that is, no eligibility requirements are set (or, alternately, requirements are discreet and do not involve any public exposure). In this case, all individuals claim for benefits.²⁰ 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 follows from the fact that $z(2) = 0$, so that 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 follows from the fact that when only deserving claimants

¹⁸ We assume that individuals decide whether to apply for welfare benefits before they observe the realization of the signal (high or low).

¹⁹ Note that the transfer, t , is assumed to be uniform (rather than means-tested, as is the case in many welfare programs). This simplifying assumption is consistent with the interpretation of the model as describing the case of a disability benefit program, where eligibility depends on having no income (in the formal labor sector), but where undeserving claimants participate in the (unobserved) informal sector.

²⁰ 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 select few as in the case of means-tested programs or programs for the disabled, significantly exacerbates the take-up problem (Currie, 2006). For a brief discussion of the take-up issue see Section 4 below.

are on welfare, the stigma cost entailed by claiming for welfare is zero; this cannot form equilibrium.

Note that an individual of type w will claim for benefits when faced with the welfare system $\langle t, \gamma \rangle$, if and only if:

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

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

$$(6) \quad t \geq s(\gamma, \bar{w}, x^*), \text{ with an equality holding when } x^* < 2.$$

3. The Optimal Policy

We next turn to formulating the optimal policy. The government seeks to attain some minimal level of utility (for all individuals) at minimal cost, by choosing the transfer (t) and the degree of exposure (γ).²¹ We simplify by assuming that exposure is costless.²² Formally, the optimal policy is a solution to the following optimization problem:

$$(7) \quad \begin{aligned} & \min_{\gamma \in [0, \bar{\gamma}], t, x} [t \cdot x] \\ & \text{s.t.} \\ & (i) \quad \underline{w} + t - s(\gamma, \underline{w}, x) \geq \hat{w}, \\ & (ii) \quad t \geq s(\gamma, \bar{w}, x), \text{ with an equality holding when } x < 2, \end{aligned}$$

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

Constraint 7(i) requires that the utility derived by all deserving claimants (the low-ability individuals) would exceed the minimal threshold, \hat{w} . Condition 7(ii) serves to determine the number of welfare claimants in equilibrium. All low-skill individuals claim benefits. In an interior equilibrium (i.e., $x < 2$), high-skill

²¹ We compare a selective welfare system with a universal one. We will consider the case of a hybrid system that is partially selective in section 3.2.

²² Allowing for costly exposure would not change the qualitative nature of our results.

individuals are indifferent between claiming and not claiming benefits. The number of high-skill individuals who do claim benefits just suffices to raise the stigma costs to a level that renders this indifference. It is straightforward to observe that constraint 7(i) is binding.²³

3.1. Comparing Selective and Universal Systems

To obtain the optimal policy, we compare the optimal universal system with the optimal selective one.²⁴ By the above considerations, the optimal universal system is simply given by $t = \hat{w} - \underline{w}$, $\gamma = 0$ and $x = 2$. That is, the transfer, t , is set at a level, which is just sufficient to ensure that the low-ability individual will attain the threshold level of utility. The cost entailed by implementing the optimal universal system, hence, is given by:

$$(8) \quad \text{cost}^{\text{universal}} = 2 \cdot (\hat{w} - \underline{w}).$$

We next turn to the more interesting case of a selective welfare system. In such a case, $\gamma > 0$, $x < 2$ and both constraints 7(i) and 7(ii) are binding. First, it is

²³ To see this, suppose, by way of contradiction, that condition 7(i) is not binding. There are two cases to consider. Consider first the case where the optimum is obtained by a universal system ($\gamma = 0$), which implies that $x=2$. In such a case, constraint 7(ii) is trivially non-binding and one can slightly reduce t without violating either of the two constraints and reduce the cost of the welfare system, thus obtaining the desired contradiction. Next, consider 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 7(ii) is binding, but then, by fixing x – the number of welfare claimants, one can slightly reduce both γ and t while maintaining constraint 7(ii) as binding and not violating condition 7(i) by continuity considerations. By reducing the cost of the welfare system, we obtain the desired contradiction.

²⁴ Universal provision of benefits has many advantages over a selective system. It enhances social solidarity, eliminates the traditional type of stigma, and has the following two political advantages: (i) it preserves the political support of middle income people for welfare programs; otherwise the welfare programs tend to shrink over time and become “poor programs for poor people,”; and (ii) selective systems tend to have an unintended marginal tax rate structure, with very high implicit marginal tax rates on low incomes due to the use of rapid phase-outs [Shaviro (1999)]. These high effective marginal tax rates are, essentially, serving as an ordeal mechanism (of the less efficient type, which also significantly harms the targeted beneficiaries), that prevents high-income earners from mimicking the low-income earners eligible for transfers. Selective systems on the other hand, have the advantage of being better targeted, which may allow the use of lower tax rates. In this paper, we focus only on the use of stigma as a potential advantage of a selective system, without discounting the importance of all other advantages and disadvantages of the two systems, which should be taken into account when designing the optimal tax and transfer system.

straightforward to verify, by virtue of the assumption that $z''(x) \leq 0$, that $\partial^2 s(\gamma, \bar{w}, x) / \partial x^2 < 0$. Thus, the expression on the right-hand side of constraint 7(ii) is strictly concave; hence, for any degree of exposure, γ , and provided that the level of transfer, t , is sufficiently small, there exist two values of x for which constraint 7(ii) holds as an equality (see figure below).²⁵ Namely, there are in general two candidate equilibria for a selective welfare system. It is straightforward to verify that only one of the two equilibria is locally stable (the one for which $\partial s / \partial x > 0$, given by x_1 in the figure). We will henceforth confine our attention to the stable equilibrium.²⁶

Fixing the degree of public exposure, γ , a (locally stable) equilibrium for a selective welfare system is given by the (unique) pair $\langle t, x \rangle$ that solves the following system of two equations given in (7):

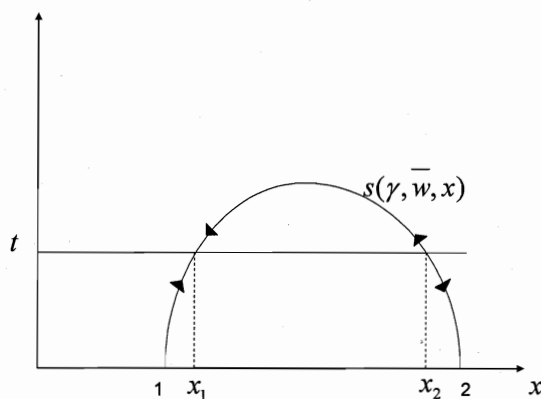
$$(9) \quad t = \hat{w} - \underline{w} + \gamma \cdot (-2q^2 + 2q) \cdot (x - 1) \cdot z(x),$$

$$(10) \quad t = \gamma \cdot (2q^2 - 2q + 1) \cdot (x - 1) \cdot z(x).$$

²⁵ This concave pattern of the stigma function is driven by two conflicting forces at play: (i) the probability of being perceived as an undeserving claimant, which rises with respect to the number of undeserving applicants; and (ii) the disutility associated with being an undeserving claimant, which decreases as the norm of claiming benefits becomes more prevalent.

²⁶ 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 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 claim for benefits.

Figure: Equilibria in a Selective Welfare System



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

$$(11) \quad \hat{w} - \underline{w} = \gamma \cdot (2q - 1)^2 \cdot (x - 1) \cdot z(x).$$

Substituting for γ from equation (11) into equation (10) yields:

$$(12) \quad t \cdot x = (\hat{w} - \underline{w}) \cdot A(q) \cdot x,$$

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

The expression on the right-hand-side of equation (12) measures the cost entailed by implementing a selective welfare system as a function of x . It follows from equation (12) that the cost increases with respect to the number of claimants. To minimize this cost, the optimal degree of exposure should be set at $\bar{\gamma}$, the highest degree that is feasible.²⁷ We thus conclude that the optimal selective system is given by the triplet $\bar{\gamma}$, $t(\bar{\gamma}, \hat{w}, q)$ and $x(\bar{\gamma}, \hat{w}, q)$, where t and x are given by the solution to equations (9) and (10) for $\gamma = \bar{\gamma}$. The cost entailed by implementing a selective system is given by:

$$(13) \quad \text{cost}^{\text{selective}} = (\hat{w} - \underline{w}) \cdot A(q) \cdot x(\bar{\gamma}, \hat{w}, q).$$

²⁷ To see this, note that as we focus on the stable equilibrium, the expression on the right-hand side of equation (11) strictly increases with respect to x . Thus, as we increase γ , x decreases. As, by assumption, increasing γ is costless, one attains the optimum by setting $\gamma = \bar{\gamma}$. Note that allowing for costly exposure would not change the qualitative nature of our results.

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

$$(14) \quad \text{cost}^{\text{selective}} < \text{cost}^{\text{universal}} \Leftrightarrow 2 / A(q) > x(\bar{\gamma}, \hat{w}, q).$$

As $1 < x(\bar{\gamma}, \hat{w}, q) < 2$, it is straightforward to verify that for a 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 $x(\bar{\gamma}, \hat{w}, q)$ strictly decreases 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 be the optimal program. Moreover, this threshold level of q strictly decreases with respect to $\bar{\gamma}$ and strictly increases with respect to \hat{w} . The latter implies that for a given $\bar{\gamma}$ 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 costs across individuals with different abilities become more pronounced. This follows from the fact that individuals can more efficiently distinguish between low- and high-ability individuals and, consequently, between more- and less- deserving welfare claimants. For a large enough q , the gains from introducing stigma costs, thereby deterring high-ability individuals from being welfare claimants, outweigh the costs associated with the need to increase the transfer (to ensure the minimal guaranteed level of well-being for low-ability individuals). As the degree of exposure ($\bar{\gamma}$) 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 cost required to support 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; hence, more individuals choose to join the welfare program. This, in turn, results in higher stigma costs required to deter potential “pretenders,” thereby forcing the government to further raise the transfers (in order to ensure that deserving individuals attain the minimal level of well-being). This amplification mechanism eventually renders the universal system (where transfers rise linearly with \hat{w}) socially desirable.

Shifting from a universal system to a selective one imposes stigma costs on participants due to the inevitable existence of undeserving claimants. 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, it is partially observable by others. In particular, the more productive an individual is, the more likely she is to be perceived as such. Hence, the more likely she is to be tagged as an undeserving claimant. This differential effect implies that higher-ability individuals face higher stigma costs from participating in welfare programs. Thus, introducing a selective system can achieve enhanced targeting. When the differential effect is strong enough (depending on the degree of exposure feasible and the amount of information possessed by individuals), a selective system would dominate.

3.2. The Optimality of a Hybrid System

Most welfare systems are neither purely selective nor completely universal. It is interesting to see whether a hybrid system, composed of a universal component and

a selective one, would be socially desirable. It turns out, indeed, that for a wide range of parameters, a hybrid (partially selective) system would be the optimal policy. Next, we illustrate this point.

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

$$(15) \quad 2/A(q) = x(\bar{\gamma}, \hat{w}, q).$$

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 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 $t^* = \hat{u} - \underline{w}$, 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 section, the optimal selective program is given by the (unique) pair $\langle t, x \rangle$ that solves the following system of two equations [similar to equations (9) and (10) above]:

$$(16) \quad t = \hat{u} - t' - \underline{w} + \bar{\gamma} \cdot (-2q^2 + 2q) \cdot (x - 1) \cdot z(x),$$

$$(17) \quad t = \bar{\gamma} \cdot (2q^2 - 2q + 1) \cdot (x - 1) \cdot z(x),$$

Note that equation (16) requires that the utility derived by a low-ability individual who participates in the selective system would be equal to $\hat{u} - t'$, so that combined with 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 section and hence are omitted), one obtains the following expression for the cost entailed by the optimal selective program:

$$(18) \quad \text{cost}^{selective} = (\hat{u} - t' - \underline{w}) \cdot A(q) \cdot x(\bar{\gamma}, \hat{u} - t', q),$$

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

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

$$(19) \quad \Delta \text{cost}^{universal} = 2 \cdot (t^* - t') = 2 \cdot (\hat{u} - t' - \underline{w}).$$

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

$$(20) \quad \begin{aligned} \text{cost}^{selective} < \Delta \text{cost}^{universal} &\Leftrightarrow (\hat{u} - t' - \underline{w}) \cdot A(q) \cdot x(\bar{\gamma}, \hat{u} - t', q) < 2 \cdot (\hat{u} - t' - \underline{w}) \\ &\Leftrightarrow 2 / A(q) > x(\bar{\gamma}, \hat{u} - t', q) \end{aligned}$$

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. Shifting entirely from a universal system to a selective one is undesirable, due to the large stigma costs entailed, while 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 imposing high stigma costs.

4. Discussion

The model outlined and analyzed in the previous sections demonstrates our key argument in favor of using statistical stigma as an ordeal mechanism to enhance the target efficiency of the welfare system. The model is made tractable by several simplifying assumptions. In the coming section we briefly discuss some of these assumptions and their implications on the qualitative nature of our results.

4.1 Traditional Stigma and the Take-up Issue

In the paper we have confined attention to the stigma attributed to individuals who are perceived to be undeserving claimants (referred to as ‘statistical stigma’). As discussed earlier on, the literature [starting from the pioneering paper by Moffitt (1983)] identified another source of stigma, which is attributed to individuals who are dependent on welfare (either through negative self image of the claimants themselves; or, via public exposure as being unable to support themselves).

One simple way to introduce traditional stigma into our model would be by letting stigma take the following extended form:

$$(21) \quad s(\gamma, w, x) = \gamma \cdot [p(w, x) \cdot z(x) + \phi(w)] + \Psi(w),$$

where traditional stigma is captured by the two functions, $\phi(w)$ and $\Psi(w)$, with $\phi(w)$ denoting the stigma component depending on the degree of public exposure (the shame of being exposed as needy); and, $\Psi(w)$ measuring the stigma attributed to negative self image (the shame of feeling needy).

Note that the major difference between the two types of stigma is reflected in the fact that the disutility associated with the traditional form is independent of the mix of deserving and undeserving claimants (captured by x). Note further, that this formulation, allows for the disutility associated with traditional stigma to depend on

the individual's earning ability, though there is no clear-cut observation as to the plausible direction of this relationship.

Traditional stigma is usually mentioned as one of the three major culprits for reduced take-up rates by deserving individuals. The other two arguments are the lack of information about the benefits, and complexity.²⁸ The problem of low take-up is often raised as an argument in favor of shifting from selective to universal transfer systems. In our model, there is no take-up problem, because all low-ability agents will take part in the selective program. This derives from the nature of statistical stigma employed, that implies that the stigma costs for low-ability agents will be lower than that of high-ability ones. If we allow for traditional stigma, some take-up problem may well arise, when the functions $\phi(w)$ and $\Psi(w)$ are non-increasing with respect to ability (which is an open empirical question). To see this, assume for simplicity that traditional stigma costs are identically and independently distributed (IID) across ability types. Although for any given level of traditional stigma, a high-ability individual will suffer more stigma costs than a low-ability agent (due to statistical stigma), there are potentially low-ability individuals who incur high enough traditional stigma costs who will suffer more disutility, all in all, than high-ability individuals who incur sufficiently low traditional stigma costs. To address the question of whether a selective program would dominate a universal program, one has to calculate the correlation between overall stigma costs and ability. If the statistical stigma forms a significant part of the total stigma costs, we expect to see an overall positive correlation between ability and stigma costs, in which case, the balance will be tilted toward the implementation of selective programs. If traditional stigma is sufficiently manifested, the balance will shift toward a universal transfer system.

²⁸ According to Currie (2006) a review of the literature suggests that the other costs associated with the take-up of social programs are more important than stigma.

Another important issue which bears on the policy implications is the extent to which traditional stigma depends on public exposure. If traditional stigma is primarily driven by negative self imagery and is but remotely sensitive to exposure, then enhancing exposure (see an elaborate discussion in the following section how this can be achieved in practice) could prove to be a powerful screening policy tool to improve the target efficiency of the welfare system.

4.2 Income Maintenance

In this paper, we assume that the government objective is to guarantee a minimum level of utility (well-being) for all individuals. Assuming alternatively that the government objective is income maintenance (ensuring a minimal level of income/consumption), the selective program will always prevail. The reason being the fact that we can use stigma to screen the low-ability individuals without the need to raise the benefit level (above the level set for the optimal universal program) in order to compensate them for the disutility suffered. Besley and Coate (1992a) demonstrate how workfare requirements can also serve to screen low-ability individuals and therefore enhance income maintenance. Note, however, that stigma is superior to workfare as an ordeal-screening device, because it entails psychic costs, which do not affect the level of income, whereas workfare requirements reduce the income level obtained by individuals.

5. Policy Implications

Assuming a given level of exposure, the government should increase statistical stigma, namely, the stigma costs associated with being perceived as potentially undeserving without affecting (or better, decreasing) the costs associated with the

traditional welfare stigma. One possible way of achieving this goal may be to publicize the names of known welfare-chiselers, thus raising the public awareness of welfare fraud.²⁹ This has no effect on traditional stigma as it directly focuses on statistical stigma.³⁰ Media coverage and general education regarding the importance of honest filing could enhance the positive effects of such publication.

The primary policy tool to affect statistical stigma is *exposure*.³¹ If we believe that traditional stigma is entirely self-characterized (Moffitt, 1983) and unrelated to public exposure, maximum exposure would be warranted. If, however, traditional stigma is increased by exposure, choosing the best policy would depend on the relative magnitude of the two types of stigma costs, which is likely to be culture-dependent and vary across geographic locations.³² In some societies, such as in the Nordic countries, the needy are viewed as entitled to government support, hence, traditional stigma is likely to be insignificant. The important question would then be the public's attitude toward welfare chiseling, which according to the World Values Survey (WVS) database is exceptionally strong. If that is indeed the case, then maximum exposure is likely to be warranted. Social norms in other societies may call for different policies.

How can exposure be achieved? Some exposure is an unintended feature of most welfare programs. The claimant usually has to attend periodic face-to-face meetings with a caseworker. These meetings are important for social work reasons,

²⁹ Increasing $z(x)$ in our model.

³⁰ Stricter enforcement would inflict costs on truly needy individuals as well, not through stigma but through their concern with being falsely accused and through greater compliance costs.

³¹ γ in our model.

³² Designing the optimal policy will therefore require the conduct and analysis of surveys. Some, however, have already been performed. See, for example, the following question taken from the World Values Survey (WVS) database, which provides questions on civic attitudes for an extensive set of countries including OECD, Eastern European, and Latin American countries: "Do you think it can always be justified, never be justified or something in between to claim government/state benefits to which you have no rights?" See Algan and Cahuc (2009) for an analysis of the design of public insurance against unemployment risk using the WVS database to explore differences across countries in the degree of moral hazard.

but they also entail exposure. Policymakers tend to view this exposure as a negative feature of the welfare program, being unaware of the potential positive aspect of stigma, the notion we name statistical stigma. And indeed the general trend is one of switching towards greater privacy. Advancements in technology, such as the ability to use the internet to download forms and fill out applications are often seen as positive developments as they minimize stigma. Similarly, one of the main arguments in support of the expansion of the EITC was its elimination of stigma [Dickert-Conlon Holtz-Eakin (2000)].

The major policy implication of our paper is simply to draw the policymaker's attention to the existence of statistical stigma and its potential contribution to enforcement efficiency. Policymakers should not necessarily tailor the programs in ways that reduce exposure. Exposure could be a desirable feature. If possible, the benefits office should be centrally located, the entrance to the building should be visible and large windows should allow passers-by to see the people waiting inside.³³

The power of exposure to passers-by may be demonstrated by the salient example of parking space reserved for the handicapped.³⁴ Passers-by can verify that the vehicle parking in the designated place belongs to a driver suffering from disability by looking at the parking permit or special license plates. Enforcement, namely, preventing non-disabled from occupying the parking spaces reserved for people with disabilities, is significantly improved by the continuous presence of

³³ The possible reduction of benefits' take-up due to a potential increase in traditional stigma because of the increased exposure can be offset by using some of the surplus attained by the increased targeting efficiency (or by reduced costs previously incurred to prevent exposure) to allow for more flexible hours for the working poor, such as adding caseworker shifts in the evening and weekends. The offices should be pleasant to visit and caseworkers should be offered the right incentives to provide professional and caring services.

³⁴ See Cope and Allred (1990) for description of the problem.

passers-by on site (contrary to official agents who can only rarely frequent the place) and the stigma attributed to pretenders.³⁵

Due to their continuous presence on site, passers-by are also likely to have information which the government is likely to lack, regarding the identity of the actual users of the car. It is illegal for people who are not disabled to use a disabled person's parking permit and/or plates to park in a reserved space. Passers-by are much more likely than official agents to be present when people get in or out of a parked car and in case they do not seem to be disabled, stigma could help in enforcement.

The most direct way to achieve exposure is publicizing the names of all recipients of certain government benefits, possibly limited to certain geographical areas due to variance in social norms, with or without explicit dollar value. The media of publication would depend on how technologically advanced the relevant society is. A government website would be ideal in the case of a technologically advanced society, and the existence of such information should be brought to the attention of the population through the general media.

This may sound disconcerting as it violates privacy. Young people are likely to find it less objectionable than older individuals. Privacy does not seem to be high on their priorities, judging by the intimate knowledge easily accessible via social networks. There is of course a difference between voluntary and mandatory provision of information, but the distinction may not be important since there is so much intimate information available on the web without required consent. Free, or inexpensive, software provides access to intimate information about people, such as their address, phone number, age, education, family members, neighbors, occupation,

³⁵ We emphasize, in the paper, the informational constraints faced by the government, and the possibility of relying on individuals' superior knowledge as a means to enhance the target efficiency of the welfare system. However, even when the government has the information, exposure to the public allows self-enforcement through social norms, thereby increasing enforcement efficiency.

and more. The income of many groups; for example, teachers and professors in many institutions, is also publicly posted. In www.zillow.com, one can find the estimated value of anyone's house, and some states allow access, through their Department of Assessments and Taxation websites, to the exact numbers. Similarly, one can find on the web, or through legal databases, the exact amounts taken in mortgages, and all criminal and civil suits and cases.

In such a reality, posting information about the identity of welfare recipients may not be exceptionally harsh. It is, after all, the taxpayers' money, and transparency with regard to the government's use of the money is generally thought to be warranted.³⁶

Another policy implication is the provision of in-kind benefits. Not merely as a means of better targeting [Nichols and Zeckhauser (1982)], as mentioned in the Introduction, but also as a way to maximize exposure. Unlike cash transfer that may involve exposure only during the process of receiving it, an in-kind benefit can also be viewed by people surrounding the recipient at the time of use. The provision of food stamps is the classic example.³⁷ Being eligible requires periodic meetings with a food stamps representative, and the actual use of the benefit is visible to others.

In 2004 an Electronic Benefit Transfer (EBT) was implemented in all states. A plastic card, similar to a bank debit card, is issued and a personal identification

³⁶ There is abundant evidence that the public wishes to control the ways in which the benefits are spent and assigns stigma to beneficiaries that use the funds in ways that are thought to be irresponsible or immoral. One recent anecdote is the story of Nadya Suleman, mother to six, who gave birth to octuplets in January 2009. The public was outraged to find out that she was receiving \$490 a month in food stamps, and that three of her first six children are disabled and receiving federal assistance. Her spending on fertility treatment was viewed to be fraudulent behavior (Los Angeles Times, 2009). Ms. Suleman also received \$168,000 in workers' compensation for back injury prior to her pregnancies. It was suggested that carrying six pregnancies, one of twins and one of octuplets, in a time span of seven years immediately after the injury, raises suspicions regarding the severity of the back injury which presumably prevented her from working.

³⁷ The program is restricted to households with incomes below 130% of the federal poverty level, or \$27,560 for a family of four. They cannot have more than \$2,000 or, in some cases, \$3,000 in assets, not including homes and, in most states, cars. The average benefit is about \$3 a day per person. Cost to the government will be around \$40 billion in 2009.

number (PIN) is assigned or chosen by the recipient to give access to the account. To preserve exposure, we suggest that the EBT card be clearly distinguishable from a debit card. This may increase traditional stigma but will clearly increase statistical stigma and preserve the exposure benefit of providing an in-kind benefit.³⁸

We do not oppose the switch from paper coupons to EBT, even though it reduces exposure, because it helps to cut back on food stamp fraud, namely, advance the same goal as statistical stigma. Paper coupons could be lost, sold or stolen. In addition, it creates an electronic record of each food stamp transaction, making it easier to identify and document instances where food benefits are exchanged for cash, drugs, or other illegal goods.³⁹

As of Oct. 1, 2008, the program name has changed to Supplemental Nutrition Assistance Program (SNAP). This seems to us to be an example of a desirable policy that decreases traditional stigma without decreasing exposure.

Public housing is another form of in-kind benefit that entails significant exposure enhancing statistical stigma. There are approximately 1.2 million households currently living in public housing units in the US. People living in public housing find it difficult to conceal the benefit, and are surrounded by people who have better information regarding their earning ability or their participation in the informal market, than is known by the government. If the local social norm is one of assigning stigma to welfare chiselers, and especially if the housing is attractive and there is a

³⁸ The introduction of EBT had no detectable effect on food stamp take-up rates and this was interpreted by Currie (2006) to suggest that stigma is not a major cause of low take-up of the food stamps program.

³⁹ See Currie (2003).

waiting list which may include relatives and acquaintances of the neighbors, statistical stigma could increase efficiency.⁴⁰

The policy implication is that public housing is superior in terms of the exposure it creates over the provision of vouchers or other housing subsidies, and that isolation of public housing projects from non-public houses is unwarranted. The environment must be mixed to prevent the creation of local social norms that may not support enforcement and the public housing should be as attractive as possible.

The exposure advantage of providing in-kind benefits in lieu of cash can be extended to additional needs for poor households. Free or subsidized public transportation can be added to the list, by providing needy individuals with an easily recognized card that allows them to use public transportation for free or at low cost, provided that they expose the card to the driver and nearby passengers.

6. Conclusion

In this paper, we discussed the effect of welfare stigma on the design and magnitude of welfare programs. The type of stigma we focused 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 differentiate between truly deserving welfare claimants and undeserving "pretenders," relying on the public's partial ability (in contrast to the government's complete lack of ability) to distinguish between the two types of welfare

⁴⁰ One downside of this form of exposure is that it cannot exclude children. This is truly unfortunate because children may experience traditional stigma, and there is no benefit in exposing them to statistical stigma, as they should not be expected to work.

claimants.⁴¹ We also establish a case for a hybrid welfare system that consists of both a universal (stigma-free) component and a selective (stigma-generating) one, as most welfare systems are, in practice.

The literature identifies another type of stigma cost, the traditional stigma, unrelated to being perceived as potentially undeserving, driven by the mere fact that welfare recipients are unable to support themselves. In putting forth our argument regarding the potential desirable feature of stigma in sharpest relief, we set aside this type of stigma cost, without discounting its importance. To the extent that traditional stigma is exacerbated by exposure,⁴² choosing the best policy would depend on the relative magnitude of the two types of stigma cost, which is likely to be culture-dependent and vary across geographic locations.

⁴¹ Welfare eligibility is often limited to individuals who cannot work. For example, the government may consider providing disability insurance benefits. The government may encounter substantial difficulties in tracing ineligible individuals who work informally but nonetheless claim benefits; that is, the government is unable to ensure that only truly deserving individuals (disabled) will receive government support. This, however, need not apply to the public that, to some extent, can observe participation in the informal sector. The 'pretending' claimants will thus suffer from a higher stigma cost than that incurred by deserving claimants who are truly unable to work.

⁴² Some of the literature (e.g., Moffitt 1983) views traditional welfare stigma as "feelings of lack of self-respect and 'negative self-characterizations' from participation in welfare," possibly unrelated to public exposure.

References

Aarts, Leo and Philip de Jong, (1996) "The Dutch Disability Program and How it Grew", in *Curing the Dutch Disease*, eds. Leo Aarts, Richard Burkhauser and Philip de Jong, (Aldershot, U.K.: Avebury, 1996).

Akerlof, George A., (1978) "The Economics of Tagging as Applied to the Optimal Income Tax," *American Economic Review* 68, 8-19.

Alesina, A., E. Glaeser, and B. Sacerdote, (2001) "Why Doesn't the United States Have a European-style Welfare State?" *Brookings Papers on Economic Activity* 2, 187-278.

Alesina, A. and G.M. Angeletos, (2005) "Fairness and Redistribution: US Versus Europe," *American Economic Review* 95, 913-935.

Algan, Yann and Pierre Cahuc. (2009). *Civic Virtue and Labor Market Institutions*. (forthcoming, *American Economic Journal: Macroeconomics*).

Besley, Timothy and Stephen Coate, (1992a) "Workfare versus Welfare Incentive Arguments for Work Requirements in Poverty-Alleviation Programs," *American Economic Review* 82 (1), 249—61.

---- (1992b). "Understanding welfare stigma: Taxpayer resentment and statistical discrimination," *Journal of Public Economics* 48, 165-184.

---- (1995). "The Design of Income Maintenance Programs" *Review of Economic Studies* 62 (2), 187—221.

Blank, Rebecca and Patricia Ruggles (1996). "When Do Women Use Aid to Families with Dependent Children and Food Stamps? The Dynamics of Eligibility versus Participation," *31 J. Human Resources* 57.

Brown, Dorothy (2007). "Race and Class Matters in Tax Policy," 107 *Colum. L. Rev.* 790.

Campolieti, Michele (2002). "Moral Hazard and Disability Insurance: On the Incidence of Hard-to-Diagnose Medical Conditions in the Canada/Quebec Pension Plan Disability Program," *Canadian Public Policy/Analyse de Politiques* 28, 419-441.

Commission of the European Communities. (1977). *The Perception of Poverty in Europe*. Brussels, March.

Cope, J. G. & Allred, L. J. (1990). Illegal parking in handicapped zones: Demographic observations and review of the literature. *Rehabilitation Psychology*, 35(4), 249-57.

Currie, Janet (2003). "U.S. Food and Nutrition Programs," in *Means-Tested Transfer Programs in the United States* (National Bureau of Economic Research Conference Report) Robert Moffitt (editor) (University Of Chicago Press).

---- (2006). "The Take Up of Social Benefits" in *Public Policy and the Income Distribution*, Alan J. Auerbach, David Card, and John M. Quigley (editors) (New York: Russell Sage).

Daponte, Beth O., Set Sanders and Lowell Taylor (1998). "Why do low-income households not use food stamps? Evidence from an experiment," *J. Hum Resources* 34:612-628.

Dickert-Conlon, S. and D. Holtz-Eakin (2000). "Employee-Based Versus Employer-Based Subsidies to Low-Wage Workers: A Public Finance Perspective." In *Finding Jobs: Work and Welfare Reform*, ed. D.E. Card and R.M. Blank, New York: Russell Sage Foundation.

Elster, J. (1989). *The Cement of Society: A Study of Social Order* (Cambridge University Press).

Farkas, Steve and Jean Robinson (1996). *The Values we Live By: What Americans Want from Welfare Reform* (New York: Public Agenda).

Feagin, J.R. (1972). "America's Welfare Stereotypes," *Social Science Quarterly* 52, 921-33.

Gallup Organization (1998). *Haves and Have-Nots: Perceptions of Fairness and Opportunity*.

Gilens, Martin (1999). *Why Americans Hate Welfare* (University of Chicago Press).

Goffman, Erwin (1963). *Stigma: Notes on the Management of Spoiled Identity*. Prentice-Hall, Englewood Cliffs, NJ.

Haider, Steven J., A. Jackowitz, and Robert F. Schoeni. (2003) "Food Stamps and the Elderly: Why is Participation So Low?" *Journal of Human Resources*, 38 (supplement): 1080-1111.

Hancock, R., et al. (2004). "The Take-up of Multiple Means-Tested Benefits by British Pensioners: Evidence from the Family Resources Survey" *Fiscal Studies* 25, 279-303.

Hecló, Hugh, (1986) "The Political Foundations of Antipoverty Policy," in Sheldon H. Danziger and Daniel H. Weinberg (eds.) *Fighting Poverty: What Works and What Doesn't* (Harvard University Press).

Jacquet, Laurence (2006). "Optimal Disability Assistance when Fraud and Stigma Matter" Queen's Economics Department Working Paper No. 1098

King, Judith (2004). "Impediments to Gainful Employment," *Social Security Journal* 66, 9-34.

Kleven, Henrik and Wojciech Kopczuk (2008) Transfer Program Complexity and the Take Up of Social Benefits, NBER Working Paper.

Lipsky, Michael (1984) "Bureaucratic Disentitlement in Social Welfare Programs" 58 Soc. Serv. Rev. 3.

London, Rebecca A. (2003) "Which TANF Applicants Are Diverted, and What Are Their Outcomes?," 77 Social Service Review 373.

Miller, David (1999). Principles of Social Justice (Harvard University Press).

Mirrlees, James A. (1971) "An Exploration in the Theory of Optimum Income Taxation", Review of Economic Studies, 38, 175-208

Moffitt, Robert (1983). "An Economic Model of Welfare Stigma," American Economic Review 75, 1023-1035.

Nichols, A., and R. Zeckhauser (1982). "Targeting Transfers Through Restrictions on Recipients," American Economic Review, 72(2), 372—377.

Orbach, Barak, (2006). "Unwelcome Benefits: Why Welfare Beneficiaries Reject Government Aid?," Law & Inequality 24: 107-150.

Pettigrew, T., (1980). "Social Psychology's Potential Contributions to an Understanding of Poverty," in Covello, V.T. (Ed.), Poverty and Public Policy.

Schenkman, Cambridge, MA.

Pudney, S. et al. (2006). "Simulating the Reform of Means-Tested Benefits with Endogenous Take-Up and Claim Costs" Oxford Bulletin of Economics, 68, 135-166.

Rainwater, L. (1974). What Money Buys. New York: Basic Books.

----- (1982) "Stigma in Income-Tested Programs," in Garfinkel, I. (Ed.), Income-Tested Transfer Programs: The Case For and Against. Academic Press, London, pp. 19-46.

Shaviro, Daniel, (1999). Effective Marginal Tax Rates on Low-Income Households, 84 Tax Notes 1191.

Super, David. 2004. "The Quiet "Welfare" Revolution: Resurrecting the Food Stamp Program in the Wake of the 1996 Welfare Law," 79 N.Y.U. L. Rev. 1271.

Weisbach, David, (2007), A Welfarist Approach to Disabilities (working paper).

Wolff, J. (2004), "Disability Among Equals", in Philosophy and Disability, K. Brownle and A. Cureton (eds.), Oxford University Pres.

Yaniv, Gideon (1997), Welfare Fraud and Welfare Stigma, Journal of Economic Psychology 18, 435-451.