Excess burden of criminality and redistribution

Dominique Demougin\textsuperscript{a,}\textsuperscript{*}, Robert Schwager\textsuperscript{b}

\textsuperscript{a}Otto-von-Guericke-Universität, Magdeburg, Germany
\textsuperscript{b}Zentrum für Europäische Wirtschaftsforschung, GmbH, Mannheim, Germany

Received \textbullet\textbullet\textbullet; accepted 17 February 2000

In the Western world, some provision for those threatened by the extremes of indigence or starvation due to circumstances beyond their control has long been accepted as a duty of the community. . . . The necessity of some such arrangement in an industrial society is unquestioned—be it only in the interest of those who require protection against acts of desperation on the part of the needy . . . F. Hayek\textsuperscript{1}

Abstract

A model where a portion of the population participates in illegal activities is developed. The propensity to participate in crimes depends on the probability and the opportunity cost of being arrested. The government can influence these variables by law enforcement expenditures and by social transfers. We analyze the cost minimizing policy mix and the incentive of wealthy agents to finance redistribution. We compare this approach with standard modelling where redistribution follows from altruistic preferences. © 2000 Elsevier Science Inc. All rights reserved.

1. Introduction

In the present paper, we examine the need for security as \textit{raison d’être} for the introduction of distributional objectives. We find that tax payers are, indeed, willing to finance transfers to reduce criminality if the safety to be implemented exceeds a critical value. Whether the

\textsuperscript{*} Corresponding author. Tel.: +49-391-671-8546; fax: +49-391-671-1218.
E-mail address: dominique.demougin@ww.uni-magdeburg.de (D. Demougin).

optimal security level reaches that pivotal threshold is shown to depend on the presence of an excess burden of criminality.

To substantiate our questioning consider one of the standard arguments advanced in literature to theoretically justify income maintenance policies. For example Wildasin\(^2\) states that “the rich are assumed to be altruistic toward the poor and, therefore, to be willing to incur tax liabilities to support redistributive transfers to the poor.” These alleged altruistic preferences are then formalized in the form of a welfare function. That function pins down a specific marginal rate of substitution between the income of wealthy and poor agents. What factors determine altruism and the ensuing marginal rate of substitution between income groups? How is this rate affected by variations in the underlying economic structure? In order to provide a possible answer to some of these questions, we introduce a model where an altruistic like behavior is founded on tax payers’ demand for safety. In our model, rich members of society consider and may, indeed, care about the income level of the less favored. They care not because of genuine altruism but because income of the poor influences their opportunity costs of a criminal activity. Seen from that perspective, transfers become one instrument to produce the public good of safety.\(^3\)

Explicitly, we model the incentives of potential criminals. We consider a world with two types of individuals distinguished by their productivity. High productivity agents pay taxes and may also become victim of a crime. Low productivity agents could receive welfare payments and may decide to become thieves. The incentive to get involved in such a crime will depend on the probability and the opportunity costs of being arrested. Governments can influence both variables by spending more or less money on law enforcement,\(^4\) and by varying the level of welfare benefits.\(^5\) In a first result, we derive the cost minimizing policy mix for an exogenously given level of public safety. We find that for a high level of criminality, it is cost effective to exclusively rely on law enforcement, and to pay no transfer. However, we also show that in order to achieve a low level of criminality, it will at some


\(^3\) See Hayek *supra* note 1. Heuristically, the argument can also be found in Pauly, M., Income redistribution as a local public good, 2 *Journal of Public Economics* 35, 58 (1973), at 38, who points out that “if poverty contributes to the incidence of crimes against property and persons, one way to reduce crime may be to redistribute income.” For an early formalization of this argument in a partial equilibrium model, see Brennan, G., Pareto desirable redistribution: The non-altruistic dimension, 14 *Public Choice* 43, 67 (1973). A more recent model highlighting the crime reducing impact of social transfers is presented by Zhang, J., The effect of welfare programs on criminal behavior: A theoretical and empirical analysis, 35 *Economic Inquiry* 120, 137 (1997). None of these papers, however, deals with the optimal choice problem of the taxpayers, which is our main focus.

\(^4\) We abstract from all considerations of private provision of security because of the well-known public good nature of safety.

\(^5\) From today’s perspective, such a motivation for the welfare state might appear quite extreme to some. However, from a historical point of view even the social work literature recognizes the relationship between the opportunity costs generated by welfare payments and criminality (e.g. Sachße, C., Von der Armutspolitik zur fachlichen Sozialarbeit, 4 *Jahrbuch der Sozialarbeit* (1981)). A prominent example is a citation by Bismarck on the introduction of the welfare state in 19th century Germany which says that the people should have “more to loose than their chains, namely a small pension” (see *id.* at 29).
point become advantageous to redistribute income. Heuristically, transfers can only be effective if there is a possibility to lose them when an agent undertakes a crime. Hence, one must first rely on law enforcement. At this stage, whether or not altruistic like behavior occurs, depends on the desired level of safety.

However, safety itself is an endogenous variable that should be determined within the model. Following standard practice, we analyze this choice from the point of view of those who finance transfers.6 This approach can readily be motivated if the proportion of productive agents is large enough. In that case the optimal choice of the productive agents also solve the median voter problem. More generally, the perspective of the wealthy is important because one would expect that it generates a lower bound solution (i.e. extending the social objective to incorporate the preferences of the poor could only increase optimal transfers). When optimizing the rich trade off the damage of criminality against the costs of security. The latter costs are given by the tax required to finance the policy mix along the expansion path and the associated excess burden. In analogy to taxes, we introduce the idea of an excess burden of crime as the difference between the gain to the offender and the damage to the victim.7 We find that this difference plays a crucial role in the analysis. When there is no such excess burden of criminality, redistribution through the tax system is never optimal.8 To use a metaphor, we are in an environment where taking the medicine makes the patient worse off than the actual illness. Specifically, to reduce the cost of criminality by 1$ using transfers would always impose a burden of more than 1$, and this for two reasons. First, because taxation causes a deadweight loss. And second, because social security cannot discriminate between potential criminals and other poor individuals. As a result, to reduce criminality via a tax-transfer scheme all the unproductive agents would have to be ‘bribed’. Consequently, in the absence of an excess burden of crimes altruistic like behavior cannot be founded on safety considerations.

However, such a deadweight loss appears very reasonable. In the case of theft, an obvious justification is that stolen goods cannot be sold at market prices. Further examples are the possibility of physical pain inflicted by an injury, the psychological suffering or even risk aversion concerns on the part of victims.9 If one of these excess burdens is large enough, it becomes optimal even for the wealthy to advocate a redistribution of income in order to increase public safety. This shows that the presence of an excess burden of crime can justify

---

6 For example, Wildasin assumes that “redistributive transfer policy is set in such a way as to maximize the welfare of the rich” (supra note 2, at 763). A similar approach is used by Pauly, supra note 3.

7 For an introduction in the concept of the excess burden of a tax see e.g. Rosen, H. Public Finance, 3rd ed., Irwin, Homewood and Boston (1991) at 305-306.

8 Thinking of Robin Hood in Sherwood Forest, it should be noted that criminality is also a form of redistribution. However, unlike the standard tax system, it only imposes a lump-sum burden.

9 For example, Anderson provides a lower bound of such a deadweight loss for the US. According to his study “the value of injuries” caused by criminality in the US amount to as much as $135 Billions per year or approximately $50 millions per 100,000; see Anderson, D., The aggregate burden of crime, XLII Journal of Law and Economics 611, 642 (1999) at 636.


11 Organisation internationale de police criminelle, Statistiques criminelles internationales 1996.
altruistic like behavior. However, the resulting trade-off between the income of the wealthy and that of the poor is not fixed by a true altruistic preference structure. Instead it reflects cost minimizing behavior of the rich where the occupational choice of the unproductive agents enters as a constraint. Accordingly, any changes in the underlying parameters, like e.g. the excess burden of crime, will affect the trade-off.

Before discussing existing literature, some rough numbers might be useful to appreciate the magnitudes involved. Table 1 summarizes some data on law enforcement and criminality for a few OECD countries. Figures on social assistance are harder to compare because the welfare systems differ across countries. Some countries guarantee a minimum subsistence level. For a single person in Denmark, this amounts to about 40% of per capita GDP. The corresponding figures in Germany and France are 33% and 25% respectively. In the US and in Switzerland, welfare is decentralized. Overall, the Swiss cantons provide a benefit level comparable to the one in Germany. In contrast, it is well known that in the USA, social assistance is substantially lower than in all these countries, in particular for young, single, healthy males. Without reading too much into these numbers, they nevertheless suggest that there is a negative correlation between the generosity of the welfare state and the amount of criminality.

Since according to our reasoning, welfare benefits contribute to reducing criminality, it may be instructive to compare the aggregate expenditures for law enforcement and social welfare payments across countries. In Germany, the latter summed up in 1998 to $26 billion per year. Given that the US has more than three times the German population, this seems high compared to about the same figure for the American AFDC program which added up to approximately $24 billions in 1995. However, Table 1 suggests that Germany recovers a large part of its social assistance payments through lower expenditures on law enforcement.\footnote{Anderson \textit{supra} note 9 estimates annual costs for police, corrections, and the legal and judicial costs of state and local criminal cases in the USA at $47, $36 and $19 billion respectively. These figures show that the lower level of law enforcement activities in Germany as displayed in table 1 translates into substantial cost savings.}

Our paper is closely related to two recent articles. Imrohoglu, Merlo and Rupert\footnote{Imrohoglu, A., A. Merlo and P. Rupert, On the political economy of income redistribution and crime, \textit{Federal Reserve Bank of Minneapolis Research Department Staff Report} 216 (1996).} investigate similar issues, though, with an entirely different methodological approach. Where we solve a stylized model analytically, they set up the problem, calibrate the relevant

<table>
<thead>
<tr>
<th></th>
<th>USA</th>
<th>D</th>
<th>F</th>
<th>CH</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police personnel</td>
<td>330.3</td>
<td>248.5</td>
<td>336.7</td>
<td>243.7</td>
<td>250.8</td>
</tr>
<tr>
<td>Prison population</td>
<td>464.9</td>
<td>55.9</td>
<td>80.1</td>
<td>53.3</td>
<td>62.1</td>
</tr>
<tr>
<td>Murder</td>
<td>7.41</td>
<td>4.32</td>
<td>4.11</td>
<td>2.81</td>
<td>4.27</td>
</tr>
<tr>
<td>Rape</td>
<td>36.10</td>
<td>7.61</td>
<td>12.39</td>
<td>4.86</td>
<td>7.39</td>
</tr>
<tr>
<td>Robbery and violent theft</td>
<td>202.44</td>
<td>82.60</td>
<td>136.95</td>
<td>31.59</td>
<td>43.42</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Statistisches Jahrbuch für die Bundesrepublik Deutschland (1998), p. 364.
\textsuperscript{b}Annuaire statistique de la France (1998), p. 967.
\textsuperscript{c}Statistisches Jahrbuch der Schweiz (1999), p. 474.
variables using US data from 1990 and then simulate some policy changes and solve for a median voter equilibrium. Of course, such a procedure is very useful because it generates some rough quantitative estimates of the relevant trade-offs. In contrast, our simple analytical framework allows us to explicitly point out the importance of the acceptable crime level and of the excess burden of criminality to understand the optimal policy mix. Benoit and Osborne\textsuperscript{14} use the same type of approach as we do and are complementary to our work. Both papers taken together employ three different instruments to reduce criminality which are (i) the severity of punishment, (ii) law enforcement and (iii) redistributive transfers. In contrast to Imrohoglu, Merlo and Rupert\textsuperscript{15} and the present analysis, Benoit and Osborne\textsuperscript{16} aggregate the latter two instruments into a single variable and study its trade-off with the severity of punishment.

More generally, our analysis is part of a broad literature initiated by Becker’s\textsuperscript{17} seminal work on the economics of crime. We follow Furlong\textsuperscript{18} by studying an equilibrium model.\textsuperscript{19} However, Furlong\textsuperscript{20} does not examine the effect of redistribution on the choice of criminal activity. This effect is touched upon in a recent paper by Chiu and Madden.\textsuperscript{21} However, there the optimal policy is not derived. Finally, Wong\textsuperscript{22} and Zhang\textsuperscript{23} also address some similar issues, though mainly from an empirical perspective. Using British time series and US cross section data, these authors confirm our main argument: the crime rate decreases if the standard of living of poor persons is raised.

Finally, our paper also contributes to a second strand of literature. Indeed, many other reasons have been advanced as to why transfers to the poor may be in the interest of taxpayers. For example, redistribution may emerge from a social contract concluded under the veil of ignorance,\textsuperscript{24} or it may provide an insurance against income shocks occurring during the life cycle.\textsuperscript{25} In this kind of argument, those who end up rich might want to avoid

\textsuperscript{15} Supra note 13.
\textsuperscript{16} Supra note 14.
\textsuperscript{18} Furlong, W., A general equilibrium model of crime commission and prevention, 34 Journal of Public Economics 78, 103 (1987).
\textsuperscript{19} For other related references, see e.g. Ehrlich, I., On the usefulness of controlling individuals: An economic analysis of rehabilitation, incapacitation and deterrence, 71 American Economic Review 307, 322 (1981); Usher, D., Police, punishment and public goods, 41 Public Finance 96, 115 (1986).
\textsuperscript{20} Supra note 18.
\textsuperscript{23} See supra footnote 3.
\textsuperscript{25} Varian, H., Redistributive taxation as social insurance, 14 Journal of Public Economics 49, 68 (1980).
taxation \textit{ex post} but redistribution increases their utility \textit{ex ante}. Moreover, Atkinson\textsuperscript{26} claims that income maintenance policies may induce the recipients to take more risk, and Grüner\textsuperscript{27} shows that transfer payments may improve the allocation of venture capital since able but poor entrepreneurs become more credit-worthy. Such an improvement in the allocation of risk presumably is also beneficial to those who finance transfers. Finally, genuine altruism cannot be ignored. Indeed, its existence can be explained by evolutionary models, e.g. Stark.\textsuperscript{28} Our approach does not contradict any of these arguments. It shows, however, that fighting criminality in itself can be a sufficient motivation for redistribution, albeit not the only one.

The remaining of the paper is organized as follows. In the next section, we present the model. In Section 3, we analyze the trade off between law enforcement and redistribution for a given level of criminality. This generates the cost function of public safety. In Section 4, we provide our main result. Section 5 offers some concluding remarks.

2. The model

As discussed in the introduction, we stylize society by a model with a public sector and two types of agents. We proceed by describing the incentives faced by potential criminals. In the aggregate we obtain a relationship expressing the level of criminality as a function of transfers and law enforcement expenditures (for brevity, we often refer to this second public activity as police but, of course, it is thought to aggregate many other policy variables like court, prison etc.). In order to finance these variables, we introduce the government budget constraint and search for the optimal policy.

The agent’s type is denoted by $\alpha \in \{0, a\}$ interpreted as productivity. The proportions of types within the population are exogenously fixed and given by $\theta$ for the non productive individuals.

2.1. Legal activity

For simplicity low productivity agents (type $\alpha = 0$) are assumed to have no labor income at all.\textsuperscript{29} However, through a welfare system non-working individuals may receive a lump sum transfer $\tau \geq 0$. For parsimony, we assume that all the agents have the same preferences which can be represented by a utility function that is linear in revenue. The utility of a law abiding type-0-person becomes [Eq. (1)]:


\textsuperscript{27}Grüner, H., Redistribution as a selection device, \textit{mimeo}, University of Bonn (1998).


\textsuperscript{29}A possible interpretation of this assumption is that the productivity of the individuals is below minimum wage.
An agent with high productivity receives a positive income net of her effort costs, which is denoted by $a$. However, in order to finance police and transfers the productive participants are subjected to a tax system. In order to raise $T$ dollars of tax revenue from a type-$a$-agent, it is necessary to impose on her a utility loss of $k(T) \geq T$. The function $k(T)$ includes the actual tax payment $T$ and the excess burden of taxation. Tax revenue is assumed to be bounded above by $\bar{T}$, reflecting the maximum of the Laffer curve. This interpretation of $k$ imposes some natural restrictions: $k(0) = 0$, $k'(0) = 1$, $k'(\bar{T}) = \infty$ and $k''(T) > 0$ for all $0 \leq T \leq \bar{T}$.

2.2. Criminal activity

Individuals may become criminally active. For simplicity, the model is designed in such a way that type-$a$-agents never choose to do so. This should not be interpreted as a better moral on the part of high productivity players. Rather it is thought to reflect different incentives. Specifically, the income of productive agents is assumed to be large enough that her opportunity cost of undertaking a criminal activity becomes prohibitive. The same argument also guarantees that, in our model, it does not pay for a type-$a$-agent to stop working in order to draw from social security.

Criminals are assumed to rob only once and either obtain $x$, $0 < x \leq a - k(T)$, if they encounter a type-$a$-agent or 0 otherwise. We denote by $p_0$ the conditional probability that a type-0-agent becomes criminal. Taking into consideration that only a low productivity agent will find it optimal to engage in a criminal activity, the probability that a randomly chosen agent is a thief becomes $p = \theta p_0$. Agents are assumed to encounter only one other person at random. Thus, $p$ also denotes the probability that an $\alpha = a$ agent will be robbed.

While $x$ gives the return of a successful crime, we allow for an excess damage imposed on the victim. Specifically, a type-$a$-person, who may be robbed with the probability $p$, sees her expected utility reduced by $d(p) \geq px$. The resulting excess burden, measured by the difference $d(p) - px$, can be justified in numerous ways. For example, we would expect that a criminal cannot sell stolen goods at market prices, thus, driving a wedge between the loss to the victim and the gain to the criminal. A further justification is the possibility of

$$v_0(\tau) = \tau.$$ (1)

30 A sufficient condition for this is:

$$a > k(\bar{T}) + \frac{1 - \theta}{\theta} \bar{T} + (1 - \theta)x + d(\theta).$$

It should be noted that conclusions would not be affected if we allowed a proportion of type-$a$-agents to become criminally active.

31 For parsimony, we have assumed that from the point of view of a criminal all potential victims appear ex ante alike. Alternatively, we could have modelled smart thieves able to identify the victim’s type ex ante. More generally, both type of thieves could coexist. Naturally, this would change some formulae, but would leave the core trade-offs unaffected. In particular, in the second part of the paper, when we ask for the willingness of productive agents to pay for transfers, introducing some smart thieves would only strengthen our results.
non-pecuniary damages to the victim either due to pain inflicted by an injury or more generally to risk aversion considerations. The function \( d \) satisfies \( d'(p) \geq x \) and \( d''(p) \geq 0 \) for all \( 0 \leq p \leq \theta \). Altogether, the utility of a productive agent becomes [Eq. (2)]:

\[
v_a(T, p) = a - k(T) - d(p).
\]  

(2)

2.3. Occupational choice

The decision process of a type-0-agent is summarized in Figure 1. Since an offender chooses a victim at random, his expected gain from crime becomes \((1 - \theta)x\). Thus, conditional on not being arrested, a thief will enjoy the expected utility \( v_{NA} = (1 - \theta)x + \tau \).

If arrested, the criminal is assumed to lose his entire monetary assets consisting of \( x \) and \( \tau \). For simplicity, we assume that these values are dissipated rather than confiscated. In addition, an arrested criminal suffers a nonmonetary utility loss from being sentenced to prison with the monetary equivalent measured by \( \sigma \), which for parsimony is taken to be independent of the success or failure of the crime.\(^{32}\) Hence, conditional on being arrested \( v_A = -\sigma \) is the expected utility of an offender. The disutility of imprisonment is assumed

\(^{32}\) Implicitly what we assume is that society punishes the intent of criminal activity. Following the well known Becker argument, supra note 17, we assume that the time spent in prison is always maximal. Of course,
to be a random variable reflecting different preferences on the part of the agents. For parsimony, \( \sigma \) is assumed to be uniformly distributed on the interval \([0, s]\).\(^{33}\)

From the point of view of a criminal, \( \rho \) denotes his probability of being arrested and found guilty through law enforcement efforts. The cost inducing \( \rho \) is \( c(\rho) \) per capita of the unproductive part of the population, hence \( \theta k(\rho) \) per capita of the overall population.\(^{34}\) We assume \( c(0) = c'(0) = 0 \), for \( \rho > 0 \), \( c' > 0 \), and \( c''(1) = +\infty \). Thus, a criminal will enjoy \( v_{NA} \) with probability \((1 - \rho)\) and \( v_A \) with probability \( \rho \). Hence, his overall expected utility becomes [Eq. (3)]:

\[
v_c = (1 - \rho)[(1 - \theta)x + \tau] - \rho \sigma.
\]

(3)

Type \( \alpha = 0 \) individuals will engage in a criminal activity if it generates a higher expected utility than respecting the law. Provided that some, but not all do so (i.e. in the case of an interior solution), we can solve for the critical agent who is just indifferent between either activity \((v_c = v_o)\). Thus, the critical \( \sigma_c \) is defined by [Eq. (4)]:

\[
(1 - \rho)(1 - \theta)x - \rho \tau = \rho \sigma_c.
\]

(4)

Because of the assumption of a uniform distribution for \( \sigma \), we conclude that\(^{35}\)

\[
p_0 = \begin{cases} 
 1 & \text{if } \frac{(1 - \rho)(1 - \theta)x - \tau}{s} > 1 \\
 0 & \text{if } \frac{(1 - \rho)(1 - \theta)x - \tau}{s} < 0 \\
 \frac{(1 - \rho)(1 - \theta)x - \tau}{s} & \text{otherwise}
\end{cases}
\]

(5)

other policies are also possible. For example in Benoit and Osborne, supra note 14, penalty is endogenous with the trade-off resulting from the risk of wrongful punishment.

\(^{33}\) From the point of view of our main conclusions, the assumption of a uniform distribution is inessential. We use it because of its simplicity and because it avoids second-order derivatives. Its main weakness is that the support must be bounded. It should be noted that in our model a raise in \( s \) shifts the distribution of \( \sigma \) according to first-order stochastic dominance which can be interpreted as making punishment more painful.

\(^{34}\) In this modelization, the per capita cost of \( \rho \) is assumed to be independent from the actual size of the criminal sector. Heuristically, police is presumed to control persons and not investigate crimes. An example of this may be the monitoring of income tax returns where a given proportion of returns are verified.

\(^{35}\) In the case that some of the thieves are smart (see Supra note 31) let \( \beta \) denote the proportion of thieves who can identify the productivity of their victim. Then the formula becomes

\[
p_0 = \frac{(1 - \rho)[1 - (1 - \beta)\theta]x - \tau}{s}.
\]
From this equation, it is immediately obvious that \( p_0 \) is decreasing in \( \rho \) and \( \tau \). This reflects that in our model police activity and welfare payments are two policy instruments that may be used to reduce crime.\(^{36}\) Law enforcement increases the risk and, thus, the expected cost of unlawful behavior, whereas redistribution\(^{37}\) increases the expected opportunity cost of losing transfers. As can be seen from the equation, this second effect can only occur when there is some level of law enforcement since otherwise transfers can never be lost.

To conclude this section, we note that our modelisation is only one of many possible ways\(^{38}\) to introduce the required degree of heterogeneity among the low productivity agents (with homogeneous individuals either all would become thief or none). A natural interpretation of our framework is that \( s \) measures the individual present value of a future punishment. The variation in \( s \) can then be taken to result from different discount rates. According to that interpretation, it is the impatient agents who become thieves, which is, indeed, a well documented empirical fact.\(^{39}\)

3. The cost of public safety

In this section, we keep the level of public safety—measured by \( p_0 \)—constant and minimize expenditures. We derive the cost function of crime deterrence and the optimal instrument mix between \( \rho \) and \( \tau \). The resulting conditions are necessary for any policy to maximize welfare. Moreover, the analysis is interesting in its own right because of the dual problem. Indeed, due to structural separation of competence many countries are organized in such a way that lower level governments with no discretion over tax revenues, have to make a choice to allocate funds to different expenditure categories.

3.1. The government budget

It is useful to give the government’s budget constraint in per capita terms. Since only type-\( a \)-agents pay taxes, the revenue per individual is \((1 - \theta)T\). From the foregoing section, per capita expenditures for police are \( \theta c(\rho) \). Finally, transfers are only paid to low produc-

---

\(^{36}\) If \( s \) is interpreted as the level of punishment—as suggested in footnote 33—then criminality is also decreasing with respect to that variable. Benoit and Osborne *supra* note 14 analyze the trade off between \( s \) and the aggregated costs associated with the other two instruments.

\(^{37}\) This redistribution effect formalizes the argument of Hayek, *supra* note 1, and Pauly, *supra* note 3, cited earlier.

\(^{38}\) There are numerous other possibilities. For example, we could have introduced individual specific moral costs (this would have marginally changed some formulae since these kind of costs would incur even if a thief is not caught) or, alternatively, assumed that criminals have different degrees of risk aversion.

\(^{39}\) See DiIulio, J. Jr., Help wanted: Economists, crime and public policy, 10 *Journal of Economic Perspectives* 1, 21 (1996) and the references therein.
tivity agents, thus, yielding an average expenditure for transfers of $\theta \tau$. Altogether, the government’s budget constraint per person becomes [Eq. (6)]:

$$ (1 - \theta)T - \theta c(\rho) - \theta \tau = 0. $$

(6)

For any given level of tax $T$, the government’s budget equation pins down a functional relationship between police and transfer which in the remaining is referred to as the isoexpenditure curve at $T$. As can be seen from Figure 2, it generates a strictly concave budget constraint defined by [Eq. (7)]:

$$ \tau = \frac{1 - \theta}{\theta} T - c(\rho) $$

(7)

with the slope

$$ \frac{d\tau}{d\rho} = -c'(\rho) \leq 0 $$

(8)

which is zero at $\rho = 0$ and strictly negative otherwise [Eq. (8)]. Higher revenues $T$ generate a parallel upward shift of the budget constraint.
3.2. Isocriminality curves

The definition of $p_0$ in Eq. (5) yields a map of strictly convex isocriminality-curves represented in Figure 3 [Eq. (9)]:

$$\tau = \frac{(1 - \theta)x}{\rho} - [sp_0 + (1 - \theta)x]$$

with slope [Eq. (10)]

$$\frac{d\tau}{dp} = -\frac{(1 - \theta)x}{\rho^2} < 0.$$  \hspace{1cm} (10)

Geometrically, reducing $p_0$ shifts the isocriminality curve outward, the intuition being that in order to reduce crime the government must either increase transfers or law enforcement or both. For $\rho = 1$, the probability of crime drops to zero even without any transfer. This is a natural property in any model with rational agents: nobody will choose to become criminally active if the probability of being arrested is 1. Hence the nocriminality-curve cuts the horizontal axis at $\rho = 1$. At the other extreme, the $p_0 = 1$-curve cuts the axis at
Heuristically, $\rho_0$ denotes the highest arrest probability such that without transfers all the low productivity agents choose to become thieves. From Eq. (11), it appears reasonable to think of $s$ as being large relative to $(1 - \theta)x$ (i.e. $\rho_0$ is small). Specifically, we assume $\rho_0 \leq \frac{(1 - \theta)x}{s + (1 - \theta)x}$. In Figure 3 the area to the left of the curve going through point $p_0$ represents the set of policies that induce $p_0 = 1$. At the other extreme, policies depicted in the area above the curve going through the point $p = 1$ generate zero criminality.

3.3. Optimal expenditure mix

In this subsection, we define the cost function of public safety, $C(p_0)$, as the minimal expenditure necessary to keep criminality to an exogenously predetermined level $p = \rho p_0$. From the minimization problem, we further obtain the optimal instrument mix. We use the definition of $p_0$ from Eq. (5). It is easily seen that for $p_0 = 1$ the cost minimizing policy mix yields $T = \rho = \tau = 0$. The intuition should be obvious: if one is ready to accept the maximal possible level of criminality one will not engage in any public safety expenditures. For all other levels of public safety, $p_0 \in (0, 1)$, we solve:

$$C(p_0) = \min_{\tau, \rho} \frac{1 - \theta}{\theta} T = c(\rho) + \tau$$

$$p_0 \geq \frac{(\rho^{-1} - 1)(1 - \theta)x - \tau}{s}$$

$$\tau \geq 0.$$  

The first inequality states that criminality must not exceed the predetermined level $p_0$. Geometrically, it implies that the set of permissible policies are to the north-east corner of the $p_0$-isocriminality curve. The second constraint [Eq. (13)] guarantees that transfers cannot become negative, i.e. that poor agents cannot be taxed. In order to solve the above minimization problem, we write the Lagrangian [Eq. (14)]

$$\Omega = c(\rho) + \tau + \lambda \left( p_0 - \frac{(\rho^{-1} - 1)(1 - \theta)x - \tau}{s} \right) + \mu \tau$$

which yields two first-order conditions

---

40 Thus, $\rho_0 < 1/2$. The assumption is used only once, in the analysis of the case without any excess burden of criminality, see proposition 3.

41 To define $C(p_0)$ as public safety is a marginal abuse of language, since it actually denotes the public costs necessary to keep criminality at $p = \theta p_0$. 

---

\[ p_0 = \frac{(1 - \theta)x}{s + (1 - \theta)x}. \]
\[ \Omega_\rho = c'(\rho) + \lambda \frac{(1 - \theta)x}{s\rho^2} = 0 \]  
\[ \Omega_\tau = 1 + \lambda \frac{1}{s} + \mu = 0 \]  
and two complementary slackness conditions

\[ \lambda \left( p_0 - \frac{(\rho^{-1} - 1)(1 - \theta)x - \tau}{s} \right) = 0 \]  
\[ \mu \tau = 0 \]  
with \( \lambda, \mu \leq 0 \) since we have a minimization problem. Using these equations [Eq. (18)], we now characterize the solution to problem (1).

First, we observe that \( \rho = 0 \). Indeed, if \( \rho = 0 \) were feasible, then from Eq. (15) \( c'(\rho) = 0 \). By assumption this would imply \( \rho = 0 \), hence \( \frac{(\rho^{-1} - 1)(1 - \theta)x - \tau}{s} = \infty \), which would violate Eq. (12). Intuitively \( \lambda = 0 \) is only feasible at the point \( \rho = \tau = 0 \), which has been ruled out.

From the second multiplier, we have two possible cases:

1. \( \mu = 0 \). In that case, we have an interior solution with

\[ c'(\rho) = \frac{(1 - \theta)x}{\rho^2} \]  
and \( \tau \) follows implicitly by Eq. (17).

2. \( \mu < 0 \). In that case, we have a corner solution with \( \tau = 0 \) and \( \rho \) follows implicitly by Eq. (17).

The optimal policy mix is summarized in the following proposition.

**Proposition 1.** Define \( \hat{\rho} \) as the solution of Eq. (19). For \( p_0 = 1 \), we have \( \rho^* = \tau^* = 0 \). For \( p_0 \in [0, 1) \) we obtain

1. if \( p_0 \leq (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \) then \( \rho^* = \hat{\rho} \) and \( \tau^* \) is implicitly defined by Eq. (17).
2. if \( p_0 > (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \) then \( \tau^* = 0 \) and \( \rho^* \) is implicitly defined by Eq. (17).

At an optimal policy mix where both instruments are used (case 1), \( \rho^* \) is defined by Eq. (19) which equalizes the marginal rate of transformation between social security and law enforcement along a budget constraint (the left handside of the equation) to the marginal rate of substitution along the isocriminality curve. Since both of these rates only depend on the arrest probability \( \rho \), it must remain constant at \( \rho^* = \hat{\rho} \) as long as one has an interior solution.

In the case of a corner solution (case 2), it can be seen from the first-order conditions of the Lagrangian that in absolute value terms the marginal rate of substitution along the isocriminality curve is larger than the marginal rate of transformation. Were we not at a corner solution, the government would want to substitute transfers for law enforcement. An
alternative way to restate this observation is that if government could tax unproductive agents \((\tau < 0)\), it would marginally do so and spend the revenue generated on police.

Figure 4 gives the expansion path which results when we vary the level of criminality \(p_0\). For the no criminality case, i.e. \(p_0 = 0\), we argue that a corner solution \((\mu < 0)\) is not possible. This follows immediately from the foregoing paragraph because at the point \(\rho = 1\), \(\tau = 0\) the marginal rate of substitution along the isocriminality curve is bounded but the marginal rate of transformation \(c'(1)\) is infinite. By continuity, the argument extends to \(p_0\) close to zero. For small shifts, we remain in an interior solution. Thus, police expenditure remains constant with \(\hat{\rho} = \hat{\rho}\) and an increase in \(p_0\) implies a reduction in transfers. This downward movement along the expansion path is bounded by one of two situations. Either \(p_0\) becomes 1 or we reach a corner solution\(^{42}\) —as in Figure 4 —where transfers are zero. Obviously, once this point is reached any further increase in criminality can only be accommodated by a reduction in law enforcement.

\(^{42}\) In figure 3, we have represented the case where for some \(p_0\) a corner solution results. This will occur if and only if \(\hat{\rho} = p_0\). This will only happen if at the point \(p_0\) a corner solution is optimal, i.e. analytically if

\[
c'(p_0) \leq \frac{(1 - \theta)\tau}{p_0}.
\]
How robust are the conclusions summarized in Figure 4? Our interpretation is that the horizontal part of the expansion path should always result. The heuristic is that transfers can only reduce crime if there is the possibility of losing them, thus, requiring an established system of law enforcement. In practice, how is the critical value of police activity determined? From the analysis, it is defined by setting transfers at zero and requiring equality between the marginal rate of substitution along the isocriminality curve and the marginal rate of transformation along the budget constraint. Consequently all the variables that affect these slopes are relevant. For an empirical application one would have to gather data on the cost function for police, , the aggregate wealth in society which influences , the distribution of that wealth stylized in the model by , and factors such as the severity of punishment or morality which we have summarized by .

The vertical part of the expansion path is an artefact of the simple specification of the model, which is due to the linear way in which transfers enter into the budget constraint and the isocriminality curve. While linearity is a natural requirement for the budget constraint, it is clearly a simplification in the case of the isocriminality curve which was introduced for parsimony. More realistically, one would expect transfers to reduce crime at a decreasing rate. This would generate an expansion path with a positive slope.

As a result from the optimization, we obtain the cost function which denotes the public expenditure of keeping criminality at . From the envelope theorem, we know that , where the second equality follows from Eq. (16). We have shown earlier that , thus, costs are strictly decreasing in criminality or equivalently increasing in public safety. As long as we are in an interior solution, the marginal cost in is . Geometrically, we are on the vertical part of the expansion path and any variation in public safety is entirely absorbed by an appropriate variation in transfers. As a result, marginal cost follows from the partial . In the case of the corner solution, the government cannot keep the marginal costs of both instruments equalized. Therefore a further increase in translates in less cost saving, i.e. . Results are summarized in Figure 5.

The last two figures are also of interest for the dual problem where a local government maximizes public safety for an exogenously given level of income. Given a level of revenue, we can read from Figure 5 the optimal and from figure 4 the resulting policy mix.

4. When would the rich desire redistribution?

In this section we complete the analysis by solving the trade-off between taxes and public safety. We do this from the perspective of a representative type-a-citizen only. Our approach is motivated by the question at hand: does the demand for public safety induce the
taxpayers to behave as if they were altruistic? Specifically, we ask under what circumstances does it become advantageous for those who finance the transfers to redistribute income.\textsuperscript{45}

The interest in this question can be further motivated from a public choice perspective. If the rich are the majority ($\theta < 1/2$) then, of course, their optimal policy choice is also the median voter equilibrium. Furthermore, as we show in the next section, in any other kind of social choice mechanism satisfying the Pareto-criterion, the optimal decision of the rich provides a lower bound for transfers.

To solve the problem of the rich, we reformulate the issue at hand in terms of the policy mix summarized in Figure 4. Specifically, when is it optimal for the rich to choose a point on the vertical part of the expansion path (the “altruistic” case) and when should they choose a policy mix on the horizontal part of that same path (the “non-altruistic” case). To answer the question, we use the cost function for public safety to define the indirect utility of a representative $a$-type agent [Eqs. (20–21)]:

$$v_a^* = \max_{T, \theta p_0} v_a(T, \theta p_0) = a - k(T) - d(\theta p_0)$$

$$C(p_0) = \frac{(1 - \theta)}{\theta} T$$

\textsuperscript{45} At this point, it is worthwhile to go back to the issue of private provision of safety which was ruled out by assumption (see supra note 4). Since in our model, all the wealthy agents are identical, their demands for safety will be the same. Thus, the choice of public safety by the representative agent would crowd out any private provision of security.
Though it is analytically feasible for the no tax solution to be optimal, we do not regard this case as particularly interesting. Such a situation would only occur if all the low productivity agents choose to be criminally active. From an empirical perspective this case does not appear convincing. Indeed, as stated by Blank, "criminal activity . . . are, even in the poorest neighborhood, the activity of a few." For that reason, we restrict attention to the case where \( p^*_0 \) is strictly below unity.

Substituting the cost of public safety in the objective function of the representative type-\( a \)-agent and maximizing with respect to \( p_0 \) yields the following condition:

\[
-k\left(\frac{\theta}{1 - \theta} - C(p_0)\right)\frac{\theta}{1 - \theta} C'(p_0) = \theta d'(\theta p_0).
\]

This equation has a straightforward interpretation. It equates the marginal damage of criminality (the right handside) to its marginal benefit from the reduction in required taxes. This benefit depends itself on the marginal cost of taxes and on the effectiveness of government revenue to affect criminality.

Since minimizing public expenditure is a necessary condition to resolving (II), we can immediately apply proposition 1 to \( p^*_0 \). Thus:

**Corollary 2**: For \( p^*_0 \in [0, 1) \), we obtain:

1. if \( p^*_0 < (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \) then \( \tau^* > 0 \),
2. if \( p^*_0 \geq (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \) then \( \tau^* = 0 \).

In this form the result is not very interesting. All it states is that transfers will be positive if the optimal level of criminality is smaller than the level represented by the isocriminality curve going through the point \( \hat{\rho} \). However, the corollary allows us to compare our approach with the standard modelling where redistribution is derived from optimization of a welfare function. The trade-off between the incomes of wealthy and poor individuals is not fixed by the form of such a function, but depends on the equilibrium behavior of individuals. Specifically, consider the definition of \( p_0^* \) in Eq. (5) which describes the occupational choice of low-productivity agents. In a situation where \( p^*_0 < (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \), one can insert

---

46 Since \( k'(\hat{T}) = \infty \), a solution will never imply \( T = \hat{T} \). On the other hand, a corner solution with \( T = 0 \) is possible because of the discontinuity of \( C(\cdot) \) at the point \( p_0 = 1 \). This corresponds to the maximum level of criminality. From figure 3, we note that the \( p_0 = 1 \)-curve is bounded away from the origin. In other words, starting at \( p_0 = 1 \), a marginal decrease in criminality requires a discrete jump in taxation. Yet the resulting benefit can only be marginal since all the other curves are differentiable. Thus, there always is a local maximum at \( T = 0 \).


48 Id. at 51.

49 It is easy to derive that the objective function is concave for all \( p_0 < 1 \). This follows from the slope and the curvature assumptions for \( k(\cdot) \) and \( d(\cdot) \) and from the observation that \( C'' \geq 0 \).
into the utility function \( v_a \) in (II) and use \( \rho^* = \hat{\rho} \) to compute the marginal rate of substitution between the tax and the transfer [Eq. (23)]:

\[
\frac{dT}{d\tau|v_a} = \frac{\theta d'}{sk} > 0.
\]

(23)

However, if \( p^*_0 \geq (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \), the wealthy agents will always choose \( \tau^* = 0 \) and the marginal rate of substitution is [Eq. (24)]

\[
\frac{dT}{d\tau|v_a} = 0.
\]

(24)

Thus the marginal willingness to pay for redistribution does not only depend on the preferences of the wealthy agents. In addition, the marginal damage of criminality \( (d') \), the excess burden of taxes \( (k') \), the incentives faced by potential criminals \( ((1 - \theta)x \text{ and } s) \), and the marginal costs of law enforcement \( (c'(\rho)) \) are important. In particular, these structural variables determine which of the cases of the corollary is relevant, and hence, whether the wealthy agents do indeed behave as if they had altruistic preferences.

In the rest of the section, we provide some intuition for the kind of economic environments which generate such altruistic like behavior. We do this by presenting several specifications for the damage function \( d(p) \) which highlight the conditions leading to a positive transfer. First, we consider the case where there is no excess burden of criminality, i.e.

\[
d(p) = px.
\]

(25)

Here [Eq. (25)], the loss of the victim is exactly as large as the gain for the offender. In this formulation, crime can be thought a ‘pure and socially costless redistribution of income’. As we now show, if this were true, it would never pay to use transfers to reduce criminality.

**Proposition 3:** If there are no excess burdens of criminality then transfers are always zero.

**Proof:** Use \( d' = x, k' > 1 \) and \( s > (1 - \theta)x \). It follows \( p^*_0 \geq (\hat{\rho}^{-1} - 1)(1 - \theta)x/s \). Q.E.D.

To understand the intuition of this result, let us initially imagine that all the unproductive population became involved with crime. It is easy to see that transfers would never be used. Indeed, as mentioned above, both transfers and stealing can be thought as one way to redistribute income, but transfers are more expensive due to the excess burden of taxes. The argument is even further reinforced when only a fragment of the unproductive population engages in crimes because transfers cannot discriminate and, thus, be targeted to potential offenders. From this we learn that redistribution will only be considered by the rich as a possible instrument when there is an excess burden of criminality. In other words, the existence of a deadweight loss of crimes is a necessary condition for altruistic like behavior to occur. It should be noted that the same conclusion would follow more generally even if we considered the costs of criminality on the less favored.

In the next two results, we examine two possible cases of such an excess burden. In the first extension, we consider situations where offenders cannot resell stolen goods at market
prices. As a result, the damage to a victim is larger than the gain to the thief. We denote with $y$, $y > x$, the loss to the victim which yields

$$d(p) = py.$$  \tag{26}$$

With this formulation [Eq. (26)], a transfer is optimal if $y$ is large enough. Specifically

**Proposition 4.** Define $\hat{y} = \frac{s}{1 - \theta} k' \left( \frac{\theta}{1 - \theta} s(\hat{p}) \right)$, then if $d(p) = py$, for all $y > \hat{y}$, the optimal transfer is strictly positive.

The proof follows immediately from Eq. (22) and is left for the reader. Heuristically, variations in $y$ leave the marginal benefit of criminality (the left handside of Eq. (22)) unchanged. However, the marginal damages of criminality increase linearly in $y$ reducing the optimal level of criminality $p^*_0$. At some point, $p^*_0$ will attain the critical value $(\hat{p}^{-1} - 1)(1 - \theta)x/s$. Any further increase in $y$ will move the optimal policy along the vertical part of the expansion path and transfers become positive.

Finally, we consider a case where the excess burden is due to risk aversion considerations on the part of the type-$a$-agents. For simplicity, we assume a mean-variance-preference structure. We abstract from other kinds of excess burden so that the expected loss is $px$ and the variance $p(1 - p)x^2$. Thus, we may write [Eq. (27)]

$$d(p) = px + r \cdot p(1 - p)x^2,$$  \tag{27}$$

where $r > 0$ measures the degree of risk aversion. As long as the probability of being a victim of a crime is less than one half, reducing $p$ will not only reduce the expected loss but also the variance. This increases the marginal cost of a crime and hence makes the optimal tax larger, and the optimal $p_0$ smaller. When risk aversion becomes large enough, again the optimal level of criminality falls below the critical value $(\hat{p}^{-1} - 1)(1 - \theta)x/s$, and the optimal transfer must be strictly positive.

**Proposition 5.** Define $\hat{r}$ as the implicit solution of [Eq. (28)]

$$\frac{s}{1 - \theta} k' \left( \frac{\theta}{1 - \theta} s(\hat{p}) \right) - \left\{ x + r \left( 1 - 2\theta \frac{(\hat{p}^{-1} - 1)(1 - \theta)x}{s} \right)x^2 \right\} = 0$$  \tag{28}$$

then if $d(p) = px + r \cdot p(1 - p)x^2$ for all $r > \hat{r}$, the optimal transfer is strictly positive.

To summarize, transfers are the more likely to be used as an instrument to reduce criminality, the higher the wedge is between the benefit of crime for the criminal and the cost of crime for the victim. If this deadweight loss of criminality is large enough, altruistic like behavior can indeed be founded on the demand for public safety on the part of taxpayers.

5. The Pareto frontier

In the foregoing section, we analyzed the case where the rich determined transfers solely for the purpose of reducing crime. Many observers would, however, argue that transfers to the poor are not exclusively paid for that reason. In the remaining, we derive two short
results. We characterize the Pareto frontier\(^{50}\) and prove the claim made in the foregoing section that the optimal transfer from the point of view of the rich is a lower bound for the solution of any welfare maximizing problem satisfying the Pareto-criterion.

To derive the Pareto frontier, we fix the utility of a representative unproductive agent at an exogenously given level \(\bar{v}_0\) and solve for the maximal feasible utility which a type-\(a\)-individual can obtain, i.e. [Eq. (29)]:

\[
\max_{\tau, p, T, \theta} v_a = a - k(T) - d(\theta p_0) \tag{III}
\]

s.t. \(v_0 = \tau \geq \bar{v}_0\) \hspace{1cm} (29)

\[
p_0 = \frac{(p^{-1} - 1)(1 - \theta)x - \tau}{s} \tag{30}
\]

\[
(1 - \theta)T = \theta\tau + \theta c(\rho). \tag{31}
\]

Feasibility requires that the level of criminality is given by Eq. (30) and that the government budget constraint [Eq. (31)] is satisfied. We followed standard practice by ignoring the

\(^{50}\) The usual economic condition for efficiency is the Pareto criterion. Considering all the feasible combinations of utility for the poor and the rich, the criterion requires that it is not possible to increase the utility of one person without reducing that of the other. The Pareto frontier is the set of all such combinations.
welfare benefits of criminality to a thieves. For parsimony, we also neglect all the potential costs of criminality to the poor.\textsuperscript{51}

The solution of (III) is represented graphically in Figure 6 for the case of an excess burden of criminality large enough that $\tau^* > 0$. It should be noted that the Pareto frontier consists only of the downward sloping part of the curve. The dotted part results when we impose $\tau = \bar{\nu}_0$.

Of course, when $\tau^* > \bar{\nu}_0$, imposing such a constraint hurts both groups in the population. The poor suffer because they get less transfers and the rich because the total costs of criminality are not minimized.

In the figure we have also drawn the social optimum for an arbitrary Paretian welfare function. As can be seen, this optimum will never be to the left of the point ($\tau^*, \nu^*$). We immediately conclude that $\tau^*$ provides a lower bound for the welfare maximizing level of the transfer.

6. Conclusion

We have shown that redistribution can in some cases be in the interest of those who finance transfers because it promotes public safety. From a cost minimizing perspective, we found that transfers across income groups are only optimal if a high level of public safety is to be achieved. Otherwise our model suggests that pure law enforcement minimizes expenditures. We also derived the optimal amount of taxes which wealthy agents are willing to pay in order to reduce criminality. We found that, when there are no excess burdens of criminality, the level of taxes chosen is so low that no transfers are ever paid. However, the assumption of no excess burden does not appear very realistic. We constructed two simple cases where an excess burden will lead the wealthy agents, for some parameters, to find redistribution optimal. Intuitively, this occurs when the excess burden is large. Thus, our paper showed that safety concerns lead to the same behavior as altruistic preferences if and only if criminality creates a substantial deadweight loss.

The conclusions of our paper were derived in a very stylized framework. A natural question is whether the deductions generalize in a more elaborate environment. Our general perception is that the core trade-offs should be fairly robust to variation of modelling specifications. For example, a natural generalization would be to introduce more than two types of productivity possibly all the way to a continuum. In contrast to our framework, the proportion of welfare recipients would have to be determined within the model. Nevertheless a prerequisite for transfer to be effective for reducing criminality is the existence of a well functioning law enforcement system. Similarly, the use of transfer can only be rational in the presence of an excess of crime.

\textsuperscript{51} Alternatively, we could introduce a damage function for the poor analogous to $d(p)$. It is easily seen that under reasonable assumptions such a modification would not alter the shape of the Pareto frontier.
Acknowledgment

The authors are grateful to Claude Fluet and Jürgen Eichberger, the participants of the 1998 conference on Public Economic Theory in Tuscaloosa, USA and the participants of the 1998 conference from the Verein für Socialpolitik in Rostock, Germany, for insightful comments.