

Assessing Deviance, Crime and
Prevention in Europe

Evaluation of safety and crime
prevention policies in Europe

Philippe ROBERT

Assessing Deviance, Crime and Prevention in Europe (CRIMPREV). Contract n° 028300. Coordination Action funded by the European Commission (FP6). Starting date of the project : July 1st 2006, duration : 36 months. Project coordinated by Centre National de la Recherche Scientifique (CNRS).

*www.crimprev.eu
contact@crimprev.eu*

Printed by the Flémal in April 2009

Lay out by JH Graphic.

N° ISBN 978 2 917565 30 8

Translated from French by Priya Vari Sen,
revised by the Author and Renée Zauberman.

Legal deposit in April 2009

**GROUPE EUROPÉEN DE RECHERCHE SUR
LES NORMATIVITÉS**

Immeuble Edison - 43, boulevard Vauban

78280 - Guyancourt - France

Tel : +33 (0)1 34 52 17 30 - Fax : +33 (0)1 34 52 17 32

www.gern-cnrs.com



*This brochure was made with the support of the
European Forum for Urban Safety
www.fesu.org – fesu@urbansecurity.org*

TABLE OF CONTENTS

I - Introduction: The context	5
1 - Workpackage on methodology and good practice	5
2 - The workshop on the evaluation of crime prevention and safety policies	8
II - The advantages and disadvantages of evaluation	10
1 - The reasons that argue in favour of evaluation	10
2 - Resisting evaluation	13
III - A variety of evaluative approaches	16
1 - A fast shrinking evaluation	17
2 - An administrative evaluation	19
3 - An example of the meta-evaluation of a safety and crime prevention policy	21
4 - An abundance of evaluations and discussions on methodology	22
IV - An ostensible evaluation without its risks	24
1 - The charms of self-evaluation	24
2 - Neutralising the external evaluator	27
V - Evaluation criteria	28
1 - Internal audit and evaluation	28
2 - Evaluation and its data	29
VI - Measuring beforehand the substance of the action being evaluated	33
VII - Evaluative know-how	37
VIII - Conclusion	42
Bibliography	45

EVALUATION OF SAFETY AND CRIME PREVENTION POLICIES IN EUROPE

Philippe ROBERT¹

I - Introduction: The context

Within the framework of the 6th PCRDT, the European Commission has financed a coordinated action *Assessing Deviance, Crime and Prevention in Europe* (CRIMPREV).

This coordinated action comprises six thematic work-packages of which one is devoted to the problems of *Methodology and good practice*.

1 - Workpackage on methodology and good practice

It entails cataloguing the most significant applications of the tools of knowledge of deviance and their uses thereof.

We thus have to:

- map the situation in Europe.
- identify the good – but also the bad – practices,
- highlight the elements of comparison within the European Union area

In the last fifty years new and powerful instruments for the study and understanding of crime have been developed. Their uniqueness lies in emancipating re-

¹ Research Director de Emeritus at Centre national de la recherche scientifique (CNRS), Centre de recherches sociologiques sur le droit et les institutions pénales (CESDIP, CNRS, ministère de la Justice, Université de Versailles Saint-Quentin).

search institutional data in which the study of crime had traditionally been embedded.

For not only have these instruments renewed – or at least partially – the scientific knowledge that we had of crime, but they are also important for the decision-making process.

Despite that, they were introduced in a very unequal manner in different European countries. Moreover, the mastery of these instruments is inequitably distributed in the sense that the number of confirmed specialists is limited; as a result, the use made of these instruments is more or less pertinent. Lastly, it often happens that the non-scientific users have not only a limited knowledge of the tools' potential, but also of their limits and the conditions of their use.

In consequence, there is place in CRIMPREV for a workpackage describing the most significant implementations of these methods in the principal European countries, and seeing how they have actually been used... in the hope of developing and disseminating the state of knowledge and the catalogue of good practices.

The responsibility for the workpackage was shared between a research centre, the *Centre de recherches sociologiques sur le droit et les institutions pénales* (CESDIP), a professor from the *Universidad central de Barcelona*, a European network of local communities dealing with safety issues, the European Forum for Urban Safety (EFUS), and a regional organisation that coordinates local safety programmes, *Città sicure*; this so-called core group was thus positioned at the interface between the academic world and the users' world.

Four methods have been selected:

- surveys among the general population on victimization and insecurity,

- surveys among the general population on self-reported crime,
- the comparison between survey data and institutional data, such as police statistics,
- evaluation of crime prevention and safety policies.

For each of these methods, the same 6-phased protocol was adopted:

- Phase 1: drawing up a report grid, the choice of a general rapporteur responsible for synthesising the information collected, and rapporteurs responsible for cataloguing the state of knowledge and its applications in countries where the methodology was sufficiently developed. Of course, it was not possible to cover all the countries in view of the lack of availability of experts. The aim was not an exhaustive coverage, but rather a reasonably representative sample of what was being done in the EU zone, especially in the main countries.
- Phase 2: a review by each rapporteur of the country or the area for which he was responsible, and dissemination of these documents.
- Phase 3: presentation of the reports and discussion during a seminar attended by the core group of the workpackage, the general and the national rapporteurs.
- Phase 4: drafting of a synthesis of the reports and discussion by the general rapporteur.
- Phase 5: validation of this document by the core group and its dissemination in the form of a fascicle in English and French.
- Phase 6: publication of the collective work in its entirety, comprising the general report and the national contributions in two volumes, one in English to be published by *VUBPress*, and the other in French by *L'Harmattan*, in the *Déviante & Société* series, under the responsibility of one of the academics responsible for the workpackage Methodology and Good Practice

The present report is devoted to the evaluation of safety and crime prevention policies.

2 - The workshop on the evaluation of crime prevention and safety policies

If, for the other workshops – surveys on victimization and insecurity, surveys on self-reported delinquency, comparison between surveys and official statistics – it was quite easy to find rapporteurs, at least in those European countries where these practices were sufficiently widespread, we encountered many difficulties in forming a team to take up the evaluation issue². It was therefore decided not to try and cover a large number of European countries but to constitute a spectrum of countries where the expression ‘evaluation of crime prevention and safety policies’ refers to very divergent practices. We were still not at the end of our troubles, for the German rapporteurs, who had agreed to participate, dropped out a few days before the seminar, after which there was no sign of them.

Ultimately, the team formed for this workshop comprised:

- Sybille Smeets and Carrol Tange (U.L. Brussels), the Belgian representatives
- Tim Hope (Keele U.), for England and Wales
- Anne Wyvekens (Centre d'études et de recherches de sciences administratives et politiques – CERSA – U. Panthéon-Assas and CNRS) from France,
- Karin Wittebrood (Sociaal en Cultureel Planbureau, SCP) from the Netherlands,
- For the general report, Philippe Robert (Centre de recherches sociologiques sur le droit et les institutions pénales – CESDIP), in place of Wolfgang Heinz (U. Konstanz) who was originally supposed to take up the assignment,

² Thus the various Scandinavian rapporteurs who were solicited, eventually did not take up our collaboration offer.

- Renée Zauberman (Centre de recherches sociologiques sur le droit et les institutions pénales – CESDIP), Amadeu Recasens i Brunet (Universidad central de Barcelona), Rosella Selmini (Città sicure, Università degli studi di Macerata), Michel Marcus (European Forum for urban Safety – EFUS) formed the core group).

In October 2007, the national rapporteurs received a letter commissioning the work from the steering committee of the workpackage and submitted their report in June 2008. After this a seminar was held at the University of Bologna and was attended by the core group, the general rapporteur and the national rapporteurs during a three-day session in July 2008, in order to present and discuss the reports, and draw cross-national lessons.

The general rapporteur thereafter prepared the present comprehensive report.

Evaluation constitutes a paradoxical subject: everyone extols its virtues but in reality they are all rather wary.

On the policy-makers' side – political and administrative – it would be delightful to have proof (with an aura of science) that *it works*, but they are always apprehensive, as Tim Hope (2009a) said, that the policy on which their success and reputation are based, may not prove to be as good as claimed.

On the academics' side, the difficulty of the method can prove a problem. There is always the dread of the ridicule, i.e. having pronounced a programme to be effective, only to later discover that it is counter-productive. Most of all, they dread having their arm twisted by the sponsors who will only accept a laudatory evaluation. Thus Sybille Smeets and Carrol Tange (2009) show that Belgian academics agreed to participate in

the evaluation bids because it was a means of penetrating areas until then closed to investigation rather than because of any great liking for evaluative research.

All in all, each protagonist, policy-maker or scholar, has perhaps a lot to gain potentially through evaluation, but also runs the risk of losing enormously. And perhaps this is what explains the real reluctance to work on a theme which however everyone talks about.

In the first section, we will first expose the advantages and disadvantages of evaluation.

We will then show the variety of ways it has been used in countries selected as typical examples.

Then we will explain how we can benefit from an ostensible evaluation without assuming the risks.

It is only afterwards that we will examine the evaluation criteria, its prerequisites and know-how.

II - The advantages and disadvantages of evaluation

1 - The reasons that argue in favour of evaluation

Strengthening the policies that are initiated with evaluative mechanisms can be an effective means of pre-empting criticism or having a ready answer for it. A policy-maker who does not bother to have his enterprise evaluated may end up with nothing to say and no counter arguments to make the day he has to face criticism. A programme whose impact is not measured will always attract a Martinson (1974) declaring that *nothing works*. And even when the criticism is without a very sound basis, or is of mediocre quality, indeed in bad faith, its effects can be devastating if, as a preemptive strategy, regular impact studies have not been carried out. Policies that could have marshalled good arguments in their favour were thus discontinued simply

because their promoters had not provided the means for proving their usefulness in time.

Hence, evaluation can first of all help to counter ill-founded criticism; but all disputes are not devoid of reason. When a policy or a programme has not undergone any evaluation for a long time, it could have gone off track so drastically without anyone noticing that, by the time anyone realised this, it would have already become so hugely detrimental and useless, and since such a long time, that it would be altogether too late to right the helm: in fact it would seem advisable to abandon it all together or change course radically³. Sherman *et al.* (1997) give a good example with their history of the *Law Enforcement Administration Agency*. As a result the impression will be left that a lot of energy and money has been expended for nothing. To go in for evaluations can help prevent such a fatal outcome.

Over and above the defensive functions, measuring the impact of a project can help to modify and adjust the plan of action. Safety and crime prevention policies cover such a vast field that their implementation is inevitably unstable and that the helm has to be constantly righted. Indeed these policies encompass all conceivable measures to curb criminal behaviour. The deterrent qualities attributed to law enforcement are far from using them up, very far. In fact, only one chapter in Lawrence Sherman's report (*et al.*, 1997⁴) to the American Congress on crime prevention⁵ is devoted to them; the others concern neighbourhood, family, school, work, criminogenic sites, and lastly the police whose actions are not only repressive, but preventive as well. We can draw up a very long list: the whole range of social relations is a storehouse to draw upon. Each of these options has its weaknesses. The very general ones

³ Example in Robert, 1994.

⁴ Extended reprint: Sherman *et al.*, 2002, 2006.

⁵A report that reviews some five hundred evaluations spread over 40 years.

run the risk of missing the target, of melting away without any notable preventive effect: this is the reproach made in the 1960s to the great supersaturating urban programmes⁶. They also run the risk of being in contradiction with the law enforcement policies conceived without any concern for synergy, as conciliating the social and the penal facets of crime control is an uneasy task. Thus the reproach often made to policing aimed at illegal immigration and banned substances is that it can mess up efforts to prevent the ghettoization of disadvantaged urban zones by creating a permanent climate of confrontation in these areas between the residents and the institutions. As for special preventive actions, it can happen that they lack the necessary backing/support structure in a very hostile environment. Thus Sherman (*et al.*, 1997) remarks that preventive programmes cannot single-handedly combat the destructive effects of urban and social segregation in urban hyper-ghettos. In such a situation, special interventions can result in exacerbated safety concerns that can have pernicious effects. In view of these dangers, safety policies rarely subscribe to one model – or at least rarely for any length of time. More generally, or at least sooner or later, they combine the models in variable proportions. Crime prevention constantly oscillates between the too general and the too specific. It is by nature an unstable public policy whose trajectory needs constant adjustments. Accompanying an initiative with an evaluation constitutes a useful tool for the policy-maker who wants to avoid the tyranny of fashionable trends and the pressure of public opinion or the self-interested advocacy of the most influential professionals.

The evaluative discourse can especially help to make the action accountable, not only vis-à-vis the political authorities who decide and finance, but also vis-à-vis the people affected by it and others living in the vicinity. This use is often neglected, perhaps because it requires

⁶ See, e.g. the report in Robert, Lascoumes, 1974.

an effort: it obviously needs a suitable language in order to be understood. It can however be highly useful in preventing antagonisms and encouraging participation. Often local authorities complain about the difficulty of involving the local residents; it should be stressed that often they do not avail of the means at hand. There should be at least one serious discussion on the basis of a reliable report on the positive and negative impact of the programmes or the expected outcome from the projects in the light of previous evaluations. This use of evaluation can also shed light on the safety expectations: the authorities obviously tend to think that their action naturally fits social expectations. It gives them the opportunity to verify if the effect of their action fulfils these expectations. To tell the truth, the problem is a more complex one. In the field of public safety as in other areas, expectations are generally diverse and often contradictory: more often than not the expectations of youth groups from neglected neighbourhoods cannot be the same as those of downtown shopkeepers or pensioners living in residential suburbs. The art of politics consists of combining divergent expectations to as great an extent as possible⁷, and making trade-offs between incompatibles, based on the specific social compromise on which, in a given society, the legitimacy of public action rests. An account of the impact of programmes in terms that can be understood by diverse audiences will make it easier to explain the choices, which form the basis of the policies, and also to highlight possible pernicious effects, which would otherwise remain unperceived.

2 - Resisting evaluation

Why then, with all its merits, does evaluation meet with such strong opposition?

⁷ Anne Wyvekens (2008) gives interesting examples of prevention schemes aimed at teaching people with heterogeneous expectations to co-exist in public places in Paris, such as the Parc de la Villette or the main railway stations.

The – little avowed – reluctance of officials, national and local, to what for them is after all a control should be taken into account. It is not that all complain about the obligation to be accountable, but if they admit to the necessity, it is often solely for the benefit of political debate, with which they are familiar. Less readily do they accept an approach, which they consider technocratic. In some countries, evaluation is even suspected of being a managerial concept that denies the specificity of public action, preferring instead modes of reasoning and action appropriate for private industry. There is some truth in this objection however, although managerial primacy favours monitoring or auditing, which chiefly verify whether the implementation of the programme corresponds to the stated objectives and calculating its cost effectiveness⁸, rendering it a mere caricature of evaluative research.

Sometimes the reluctance of the authorities is due to a less noble concern than that of protecting the sphere of political debate from managerial interference. For example, keeping some leeway may be a concern: thus, in the general crime prevention schemes, local authorities find in funding programmes accommodating ways for promoting social action without specifically targeting crime prevention. Or it may simply be the reluctance to account for their actions, the conviction that the policy maker knows what is useful and good for society, thus rendering all verification superfluous.

The evaluation can also be missing, not because of the reluctance of the authorities for one reason or another, but because their expectations are too divergent to leave the evaluators sufficient margin: torn between the strong assumptions of the national political forces,

⁸ Examples in Leigh (2000) – regarding the obligation of a 2% improvement of efficiency per annum in the Crime and Disorder Act – and Braga et al. (1999), who stress the importance of Value for Money estimation, in the reform programme of the New Jersey City police.

the contradictory demands of the Department of Finance, the police, social workers, teachers, doctors, prosecutors, judges, local elected representatives, associations... they are not able to gain sufficient autonomy for serious work. Ultimately, nothing is evaluated and the trade-offs are made only on the basis of the ongoing power struggle between the different protagonists.

Some measures resist evaluation more than others. First, strictly speaking, an entire policy can only be evaluated by a sort of meta-evaluation, which is the sum of the evaluations of the different programmes that compose it. After the Dutch Court of Audit had criticised the cost-effectiveness of the country's crime prevention policies, L. Van Noije and Karin Wittebrood (2008) had to review 149 evaluations to evaluate the policy *Naar een veiliger samenleving* (Towards a Safer Society), launched in 2002 by the first Balkenende government. In addition, it has often been maintained that social prevention measures are more difficult to evaluate than situational crime prevention measures. Thus Karin Wittebrood (2009) states that programmes that claim to tackle the supposed causes of deviance such as missing school, the quality of parenting, alcoholism, are unevaluable in so far as this causal relation remains hypothetical⁹. The difficulty probably arises from the fact that social policies only indirectly endeavour to reduce crime and insecurity, i.e. they try to contribute to such effects through their direct impact on the living conditions of the target populations. Lastly, surprisingly, law enforcement is much less subject to evaluation than prevention¹⁰, although it has greater

⁹ For all that, Hope (2009a) recalls that the New Deal for Communities Programme (NDC) which pursues reduction of unemployment, of crime, improvement in the fields of health and employment, regeneration of housing and environment in 39 particularly underprivileged localities, was the subject of a national evaluation programme.

¹⁰ One of the merits of the Dutch meta-evaluation described by Wittebrood (2009) consists precisely of not skipping the penal sphere. See however Mackenzie, 2006.

need as it depends overwhelmingly on unproved assertions¹¹ and the evaluation faces fewer difficulties for it entails individual and mandatory participation of the target-population whereas prevention seeks to mobilise participation which is collective and voluntary¹².

Faced with such a contradictory situation we will first begin by reviewing the evaluation methods adopted for prevention and safety policies in some European countries.

III - A variety of evaluative approaches

Rather than cover a large number of European countries, we have constituted a sort of sample of the very varied practices and even conceptions of the evaluation of safety and crime prevention policies.

In Belgium the evaluative effort – which is linked to safety contracts between the Federal State and the municipalities – seems to have lost momentum in the last few years.

France offers an example of an ‘official’ evaluation generally undertaken by the administration.

The Netherlands has chosen a meta-evaluation that examines all the available country-wide data on the basis of ‘quasi-experimental’ scientific canons of Anglo-Saxon origin.

Lastly, England and Wales present old and systematic evaluative techniques which cover the entire range of prevention policies, making it possible to address complex methodological problems.

¹¹ As Lawrence Sherman said in 1997 about the massive recourse to imprisonment.

¹² As Tim Hope points out (2009a).

1 - A fast shrinking evaluation¹³

The safety contracts signed between the Federal Ministry of the Interior and the municipalities were accompanied by evaluations to decide the continuation, extension or reduction of the funds released by the Central government. The totality of these ‘evaluations’ are called ‘summative’ in the sense that their aim is to make an assessment of the contracts between the State and a municipality in order to decide whether to continue it or not.

Some evaluations – termed ‘administrative’ – are in the form of tables giving an overview of the number of contracts and the localities involved, budgetary allocations, the mechanisms retained, and sometimes police statistics or data taken from the Security Monitor¹⁴.

Others – termed internal – have to be carried out by agents paid by the municipality benefiting from a security contract. But, since 2005, these evaluations have become (preliminary) safety audits – mandatory since 2007 – carried out on the basis of a methodological guide developed by the Ministry of the Interior.

Along with this, a research programme was initiated in 1985, again by the Ministry of the Interior, firstly on the police. From the commencement of the following decade, it was extended to safety and crime prevention policies in order to determine the weaknesses in the programmes and to contribute to their improvement.

¹³ Based on the report on the Belgium example presented by Sybille Smeets and Carrol Tange (2009 and ref. cit.) at the Bologna seminar.

¹⁴ A questionnaire prepared by the police and administered over the phone by a private firm to a random sample of some 40,000 interviewees. The questions were on neighbourhood problems, victimisation, insecurity, contacts with the police and opinions about its functioning. It has been administered 6 times since 1997.

This was called ‘formative’ evaluation. During the entire last decade of the 20th century, academic research teams thus studied the procedures and functioning of the programmes on the local police, the community policing, rapprochement between the police and the citizens, assistance to victims, re-entry into the labour force, hooliganism, school dropouts, and project coordination... This involved mainly qualitative research, which did not specifically plan to measure the impact on crime and insecurity.

From 1999, the calls for tender responded to the need of greater operationality. They were addressed to government bodies and private structures. The amount contracted for carrying out the research and its duration was considerably reduced. Far from seeking to evaluate the impact of the programmes and the policies, these studies are more like an inventory of resources, or suggested methods of management, of policy planning and of data analyses. Nevertheless, the results of these works are not really heeded by their sponsors, except if their angle is to ‘know thyself in order to best justify thyself’.

Until then, public policies tried to compromise between a conservative trend more in favour of reinforcing the punitive measures, and a social-democrat trend, which opted for local safety contracts. Similarly, the evaluations made a sort of compromise between an assessment of the outcome of the programme and a qualitative study of the discourse and personal experience of the actors. In recent years, the emphasis has been on streamlining the resources devoted to safety and the concentration of evaluation funds on straightforward management studies in which the academic sphere is increasingly sidelined.

2 - An administrative evaluation¹⁵

Since the Rocard government of the late 1980s, the evaluation of public policies has become a totally official activity. A statutory instrument was even promulgated some ten years later: Decree 98-1048 of November, 18, 1998 on the evaluation of public policies. This activity seems to be generally conducted by the concerned departments, mostly with the support of the inspectorate which exists in each administrative branch. The most prestigious among them, the *Inspection Générale des Finances* (IGF), holds a pre-eminent place, generally in association with the inspection belonging to the relevant ministry. One of the permanent tasks of the *Cour des Comptes*, the French court of audit is also the evaluation of public policies. The place of these two bodies underlines the highly ‘financial’ bent of the official conception of evaluation ‘à la française’: primarily it involves verifying the proper use of public monies.

With regard to safety and crime prevention policies, however, we can only cite a limited number of examples. In 1993, the *conseiller d’État*, Jean-Michel Belorgey, had presided the evaluation mission of the *politique de la Ville*¹⁶ but without giving a very prominent place to crime prevention programmes which had been clubbed with this public policies package from the late 1980s. This report was a follow up to the review of Neighbourhood Social Development local contracts – the ancestor to the *politique de la Ville* – which was

¹⁵ Based on the report presented at the Bologna seminar on the French example by Anne Wyvekens (2009 and ref. cit).

¹⁶ A set of public policies addressing impoverished urban areas. It should be remarked that the task force in charge with this mission was not only composed of civil servants: for its better part, its preliminary work had been conducted by academics, mainly Jacques Donzelot who has drawn upon this expertise to publish a book with Philippe Estèbe (1993).

completed in 1988 under the direction of François Lévy, an *ingénieur général des Ponts et Chaussées* (Engineer General of Bridges and Highways).

This quasi administrative monopoly of policy evaluation, especially in crime prevention and safety matters, is seemingly accompanied on the part of policy-makers and officials by a solid mistrust of academia, who they feel are too quick to criticise public programmes.

For their part, academic circles do not seem to be particularly keen on this evaluation activity. Only political scientists specialising in the analysis of public policies show some interest in it¹⁷, although not all of them. This outlook does not really retain the attention of those who use a bottom-up approach to study public policies, i.e. interactions between the street-level administrative officer and the client, or from the point of view of the latter. There exists a sort of scholarly consensus which estimates that what is undertaken by the administration in the name of 'evaluation' constitutes merely an internal audit, and that there is practically no genuine evaluation of public policies done in France, especially on safety and crime prevention.

However, the *Institut national de la statistique et des études économiques* (INSEE, National Institute for Statistics and Economic Studies) has conducted a study on the trend in the relative situation of neighbourhoods where the *politique de la Ville* was implemented between the 1990 and 1999 censuses. This work by government statisticians is what best approximates an evaluative study.

For the most recent period, among the many national observatories operating in the field of prevention and safety policies, those dealing with disadvantaged urban zones and above all drugs and addiction have

¹⁷ See, e.g., Lascoumes, Setbon, 1995; Kessler, Lascoumes, Setbon, Thoenig, 1998.

published reports that are clearly in line with an evaluative perspective.

We can still cite some specific studies, for example on the use of CCTV in high schools of the Île-de-France region (Paris and suburbs) but it is only to point out that their dissemination has been seriously curtailed by the instructions of the sponsors.

Lastly, mention should be made of a certain number of studies, evaluative in nature, on the penal career of convicts sentenced to various types of penalties, after their release.

3 - An example of the meta-evaluation of a safety and crime prevention policy¹⁸

Queries by the Dutch Court of Audit regarding the performance of a crime prevention and safety programme, initiated in 2002, have led to a meta-evaluation in three phases by the Social and Cultural Planning Office/Institute of Social Research of the Netherlands (*Sociaal en Cultureel Planbureau, SCP*) and assigned to Van Noije and Wittebrood.

These authors first collected some 150 evaluations on various aspects of this policy.

They then selected those that they thought conformed to the evaluative principle defined in the classic report for the United States Congress prepared under Lawrence Sherman's direction, and also with the criteria laid down by the *Campbell Collaboration Crime and Justice Group* for promoting the development of a quasi-experimental evaluation in the field of crime and criminal justice.

¹⁸ Based on the report presented at the Bologna seminar on the Dutch case presented by Karin Wittebrood (2009a and ref. cit).

They ultimately used the studies selected to review the different chapters of the government policy under scrutiny – law enforcement, developmental crime prevention, situational prevention, and lastly, systemic measures – while emphasising that all the studies necessary to review several aspects of the programme initiated in 2002 were not always available, and that it was necessary to encourage a still more systematic development of quasi-experimental evaluation.

It is not the first time that the Dutch government has thus asked for a systematic review of the available evaluations¹⁹. The raw material has been provided thanks to the old habit of reserving 10% of the amount allocated to crime prevention and safety programmes for evaluation. Such a routine tends to show a certain practice of meta-evaluations, although, each time, the reviewers deplored the extremely inconsistent quality of the evaluations carried out.

4 - An abundance of evaluations and discussions on methodology²⁰

In England and Wales, since the 1980s, evaluation of the crime prevention and safety programmes has been systematically carried out. What is more, since its return to power in 1997, the Labour party has combined its policy of crime and disorder reduction with an *Evidence-based policy and practices (EBPP)*, which primarily rests on the accumulation of expert knowledge acquired from the systematic review of evaluations. In undertaking the latter, the standardised criteria developed in the aftermath of Sherman's work at the University of Maryland and that developed by the group

¹⁹ See Junger-Tas, 1989; Polder, Van Vlaardingen, 1992; Soetenhorst de Savornin Lohman, 1997; Van Dijk, 1993.

²⁰ Based on the report presented at the Bologna seminar on the Anglo-Welsh example by Tim Hope (2009a and ref.cit.).

pursuing the work initiated by Donald Campbell, carry more and more weight.

We can thus count the evaluations carried out:

- *Home Office Crime Prevention Feasibility Study* based on research using the case study method;
- *Neighbourhood Watch Evaluation* that has adopted a quasi-experimental approach on the basis of survey data;
- *Community Policing Fear Reduction*, also of a quasi-experimental bent;
- *Priority Estates Project*, quasi-experimental too;
- *Kirkholt Burglary Reduction Project* that focuses on an assessment of the effectiveness of programmes aimed at repeat burglaries;
- *Safer Cities* programme, which includes the sophisticated use of micro-econometric methods by Ekblom *et al.* (1966), but also another one using qualitative techniques (Tilley, 1993b);
- *Design Improvement Controlled Experiment (DICE)*, using a range of techniques including survey data;
- *Communities that Care (CtC) Demonstration Projects* for the Joseph Rowntree Foundation by a research group from Sheffield University;
- *New Deal for Communities Programme (NDMC)*, employing a battery of indicators and analytical approaches;
- *Crime Reduction Programme (CRP)*, combining some 80 evaluative studies based on the reduction of burglaries and remote surveillance through CCTV;
- Reassurance Policing.

This abundance of works using a variety of methods and data enables a detailed discussion of evaluative methodology. This puts the British rapporteur in a position to identify the flaws in an evaluation limited to the 'black box' model, discuss the merits of Pawson and Tilley's 'realist experiment' evaluation, show how the

‘quasi-experimental’ approach itself (Farrington, 2003) comes up against difficulties, to resolve which he proposes borrowing elements from micro-econometrics.

The systematic use of evaluation in piloting prevention and safety policies can facilitate the discussion on the consequences of such an extensive mobilisation of scientific resources and their stowing to the goals of public policies. In this regard, Hope (2009a, b) suggests that such a situation could paradoxically lead, not to the ‘scientification’ of the politics, but to the ‘politicisation’ of science, if it is combined with a compelling aversion to critical results. There can be no contribution of science to politics except if each maintains its autonomy in relation to the other, and therefore its own qualities.

That the evaluation of crime prevention and safety policy refers to a series of highly contrasting practices in European countries, which in other respects are quite similar to each other, highlights the difficulty of outlining a synthetic review. Such an endeavour should be able to deal both with the problems engendered by the lack of evaluation, and with those caused by its over-utilisation.

IV - An ostensible evaluation without its risks

The temptation is great to avail the benefits of evaluation without subjecting oneself to risks. There are two ways of doing this... The first consists of evaluating oneself, the second is to control the external evaluator really well so that he is practically forced into arriving at only positive conclusions.

1 - The charms of self-evaluation

Self-evaluation is the simplest.

Thus, according to Smeets and Tange (2009), the organisational reform of the Belgian police meant that

the policing projects of the safety contracts²¹ were reserved for an internal evaluation undertaken by this very department. Wyvekens (2009) gives an even more significant example of ‘the evaluation of public policies’ in France, an administrative activity governed by statutory instruments²², whose responsibility lies with the inspectorates attached to the various departments. In reality the reviewer is too close to the policy implementation to produce a genuine evaluation and these so-called ‘evaluations’ are, at best, internal audits. Sometimes, they are not even that: Roché (2005) cites the example of the ‘evaluation’ of the community policing initiatives by inspectors attached to the department of the Interior: he observes merely *a long list of subjective judgements*.

With regard to the denunciation of self-evaluation, the specialists concur and sometimes in rather harsh terms²³.

From the inside, the risk is too great of mixing up the output of a programme with its effects on a population or a territory, for a whole series of wrong reasons. From this position, the access to internal data is so easy that the evaluator is tempted to use it as such²⁴ without meticulously examining whether it really measures what has been done or the impact that this action could have had²⁵.

²¹ Which later became a strategic plan for safety and crime prevention.

²² Decree no. 98-1048 of November 18, 1988 relative to the evaluation of public policies.

²³ See, e.g. Brodeur, 1998b.

²⁴ Hence, according to Smeets and Tange (2009), the Federal Ministry of the Interior obliges holders of security contracts to use the department’s internal statistics ‘in order to be objective’ (sic).

²⁵ They could possibly even be presented so as to give the impression of a positive result.

From the inside, it is difficult to get away from the logic of action: if the objectives were not implemented or if they were incorrectly implemented, that can easily be observed; but if they were, it will be difficult to envisage that the effect has not necessarily reached the target or that it differs from what was planned. In any case, if the internal auditor manages to discern the intended effects, he is in the worst possible position to see the unexpected, side- or displacement effects.

In practical terms, this requirement has two consequences. Evaluation can be asked neither from an organisation entrusted with the programme's conception or its implementation, nor from an organisation entrusted with the monitoring of a public programme, for example an inspection or an audit body.²⁶

Self-evaluation only produces illusory results, a surface evaluation. In her review of evaluations available in the Netherlands, Karin Wittebrood (2009), concludes that none of those carried out internally by the sponsors of the prevention and safety programmes have even attained the minimum level of validity.

To keep up appearances, a substitute is to call in private consultants. Smeets and Tange (2009), as also Wyvekens (2009), give multiple examples of the preference for this solution both in Belgium and France. By and large unrestricted by the principles of evaluative research, they are as quick as their sponsors would like them to be and their conclusions are not upsetting. In the Netherlands, Wittebrood (2009) considers their technical level at times acceptable, at times inadequate, but critical conclusions are never delivered.

²⁶ Thus the evaluation scheme of the Chicago police reform was entrusted to a consortium of local university departments sponsored and financed by the State of Illinois. It was completely independent of the municipality and the police corps, who are in charge of conceiving and implementing the CAPS programme (Skogan, Hartnett, 1997a, b).

2 - Neutralising the external evaluator

It can easily be comprehended that self-evaluation (or the recourse to obliging private consultants) has all the appearance of an evaluation without incurring its risks. It seems more of a paradox to locate a comparable risk in an excessive 'evaluation'.

However, the systematic integration of an evaluative project in all crime prevention and safety programmes can lead to a type of domestication of the external evaluator, who becomes party to government contracts at the risk of losing some of his independence, somewhat like the journalist who is embedded in an army unit in action.

There is no full guarantee that even the strict standardisation of evaluative research operations suffice to ensure against this danger. Methodological techniques can constitute an illusory defence that are brandished all the more readily as the other criteria of scientific production like peer reviews and the publication of procedures and results in an easily replicable format are sacrificed.

In this case, far from achieving the 'scientification' of political action considered by the initiator of the procedure, Donald T. Campbell (1969), as being the criterion of a good 'open' society, what is actually achieved is 'politicisation' of scientific activity.

It is probably this risk, rather than the difficulties of the method, that deters many researchers from taking up evaluation work.

But in which case, the recourse to evaluation will not really bring any added value, which presupposes accepting the risk of a negative conclusion by the evaluator, and also signifies that the latter himself is prepared to conclude negatively as well as positively.

Hence, recognized evaluators (e.g. Hope, 2009b) insist on a punctilious respect for autonomy and deontology peculiar to the scientific world in a discipline where the political stakes are extremely high, as insecurity has become less a problem needing a solution than a lever of governance (governing through crime).

V - Evaluation criteria

The weakness of all these shortcuts seeking the benefits of evaluation without incurring risks stem from a mistaken idea of evaluation.

1 - Internal audit and evaluation

Comparing the objectives with their implementation, determining the costs and comparing them with the initial provisions, describing the output, examining it in relation to the objectives, all these steps resemble an audit, a cost-effectiveness calculation (value for money) or programme monitoring, but they do not constitute an evaluation²⁷.

Confining the assessment to measuring what has been done is not an evaluation. Evaluation really begins with the determination of the outcome. An example²⁸ will highlight the difference. The setting up of twenty CCTV cameras in the streets of a town constitutes an output, not an outcome: this is what has been done. On the other hand, achieving a 20% reduction in crime on the streets and minimising fear definitely constitutes re-

²⁷ See e.g. Sherman et al. (1997): a description of the implementation, its monitoring do not compare with an evaluation; see also Brodeur, 1998b or McKenzie, 2006, 342.

²⁸ Borrowed from the manual edited by Hough and Tilley (1998) for the Home Office for the purpose of guiding the local authorities in their implementation of the 1998 Crime and Disorder Act.

sults. Verifying that what had been planned has indeed been achieved is not yet an evaluation, even if this is a required preliminary step. *Evaluation begins with seeking, not what has been done, but the effect that this action has had on the target.* Often, however, it is wrongly believed that an evaluation has been achieved, whereas the substantiation was not pursued that far, but was limited to recording the outputs and not the outcome.

This can be practised from a managerial perspective, which stresses that implementation must meet the objectives, also that expenses should be monitored, and which offers different techniques of verification. Paradoxically, this approach can also result from a strong administrative tradition whereby a central agency has the necessary resources to carry out inspections and verify whether the implementations correspond to the intentions of the politicians.

2 - Evaluation and its data

This clarification on the very nature of the exercise affects its materials: as a general rule, the internal data recorded by the concerned body is not enough for an evaluation²⁹.

The reason is simple: those records only catch events that came across institutional action. Having a burglary or a mugging figure in police statistics depends not only on the occurrence of a victimisation, but also on the propensity of the victim to lodge a complaint and that of the police to record it in a format that allows reckoning.

One example will illustrate the difficulty: take an increase in the police recorded crime. Should this lead to the conclusion that the crime situation has worsened?

²⁹ See e.g. Brodeur, 1998b.

Not necessarily. On the contrary, surveys can actually reveal that there has been a decrease in the number of victimisations. It is the propensity of the victims to lodge complaints that has gone up³⁰ or their recording that has become more systematic³¹. Relying on official records alone could lead to a mistaken interpretation of the observed trend.

Moreover, counting does not always help to appreciate the prevalence of victimisation. For example, the increase in the number of recorded offenses can be due to the enlargement of the affected population or, on the contrary, to an aggravation of incidence – the number of victims has not increased but each victim has undergone repeat experiences. To appreciate the impact of a crime prevention programme, these two hypotheses have necessarily to be distinguished, diluted victimisation or concentrated victimisation. Often, only a survey will determine this.

Lastly, the effect of victimisation depends greatly on the manner in which it is experienced. One of the French surveys³² came across victims who were not very traumatised – their mishap was due to their lifestyle and a slight change in it could easily prevent a repeat, for example not roaming in certain areas at certain times. Whereas others took it very badly, for although the incidents were only marginally more serious, the victimisation was linked to their living conditions: for example, they would have had to move from their present residence in order to be safe and they did not have the necessary resources. To evaluate the impact of crime prevention programmes involves distinguishing between these two types of experiences. Here too, simple institutional data is not enough, a social survey is imperative.

³⁰ This example is given by Wittebrood (2009) in connection with the Dutch programme. Similarly Sherman's (et al., 1997) example described the American experience.

³¹ Example in Hough, Mayhew, 2004.

³² Zauberman, Robert, Pottier, 2000.

Tilley (1995) also notes that institutional data can be presented in a biased manner – it is quite understandable but nonetheless misleading – so as to give the impression that the programme is effective, even if the results are indecisive.

But using pre-existing institutional data is generally a cheap and (relatively) quick solution, whereas doing a survey is both costly and time consuming³³. So what is to be done? Firstly, other pre-existing data from other sectors can be resorted to³⁴ allowing cross-checking: the unpaid rents in public housing, statistics on school incidents or truancy... It is then possible to pre-position the surveys, for example by carrying out, within the framework of a regional crime observatory, regular surveys on victimisation, insecurity, self-reported crime. Even if certain areas have to be oversampled in order to gain specific knowledge about them, the cost will be less than starting a survey from zero, and above all it will be quicker. Also it would be better to evaluate specific programmes and sanction the funds needed for a useful survey, rather than pretend to evaluate everything without sufficient data. It is here that the stipulation setting aside a percentage of the funds allotted to crime prevention policies for evaluation, as is the custom in the Netherlands, is very useful. Lawrence Sherman (*et al.*, 1997) says it explicitly: they must make it feasible to carry out costly surveys which are required for a serious evaluation.

Lastly, one should not wait for the programme to be implemented, or what is worse, completed before thinking about which data will be useful for its evaluation. Some data should be collected before starting in order to compare the *ex ante* and *ex post* situations.

³³ Ekblom, 1998; Hough, Tilley, 1998.

³⁴ E.g. Hough, Tilley, 1998; compare Lagrange, Sina, 1998, although they deal with preliminary audits and consider that an evaluation should resort to other data.

On the subject of data, one question is also whether quantitative data is alone valid for evaluative research. *A priori*, nothing counter-indicates that qualitative data should be used. Recently, at the Marc Bloch Centre in Berlin, in my conclusions at a political scientists' conference assessing public policies via the interactions of street level bureaucrats with their clients, I remarked that some of the ethnographic observations could have provided good evaluative material. Moreover, we know that Tilley (1993a, b) has used qualitative techniques in the evaluation of portions of the *Safer Cities* programme. Pawson and Tilley (1994, 1997) have argued in favour of a 'realist' evaluation that resorts to qualitative analyses, a view which has led to a disagreement with the promoters of the *Campbell Collaboration Crime and Justice Group*, and in particular David Farrington (2003). In truth, the discussion goes beyond the mere status of data: it is more concerned with the review of a 'quasi-experimental' method, which Pawson and Tilley criticise for focusing only on what works. These authors regard as evaluable only what has worked in a given context and in view of this context... on the basis of rather ethnographic observations. Whatever the outcome of this quarrel, it is not certain that it condemns the use of qualitative techniques in the evaluation³⁵.

In any case, measuring the impact of a programme entails not confining the approach to a simple audit or a management control, separating the evaluation from the running of the programme, and lastly, seeking the resources to generate data external to the programme under scrutiny.

³⁵ All M. Q. Patton's work (see especially 1980) constitutes a forceful plea in favour of a qualitative approach in evaluation. He defines the latter as: determining how a programme affects people (328).

VI - Measuring beforehand the substance of the action being evaluated

For all that, if the estimation of what has been done cannot be considered an evaluation, it nevertheless constitutes a mandatory first step in order to avoid measuring the impact of what did not really take place.

One should therefore start by determining the substance of the object whose impact we are examining. Few spheres of public action are more subject to announcement effects than crime: there are talks and promises, but all these statements do not constitute faithful descriptions of what is really going to be undertaken. Failure due to the ineffectiveness of a programme has to be differentiated from that attributable to a nonexistent or uncompleted implementation. It is generally agreed that the preliminary phase of an evaluation should take into account the *objectives*, the *inputs*, the *implementation* and the *outputs*.

It is the former that poses the biggest problem: generally, the boat is overloaded and a multitude of objectives publicised, sometimes without any clear relationship with the action taken³⁶. In case of multiple partnerships, which is usually the situation, these aims can differ from one partner to another and sometimes they even contradict themselves³⁷. There are some who

³⁶ Thus Brodeur (2000) notes the distance that exists between the prolixity of the discourse on Community Policing and the relative inadequacy of the concrete steps taken for its implementation.

³⁷ To the extent of challenging any evaluation that wishes to focus on this contradictory multiplicity: Crawford, 1998b. Thus Breen (2002) and his associates, in their work on the evaluation of the justice system, note the lack of consensus over its objectives; they set about determining some by themselves, only to be disclaimed by their sponsors who, in their implementation 'forgot' those of these objectives that could have turned embarrassing, like impartiality or equality (Beauvallet, 2009).

feel it is necessary to distinguish declared from latent aims. This entails the risk of setting off on an exhausting race to determine the intentions. Would the task be easier when moving downwards from global policies to individual programmes? I have come to the conclusion that it is not worth dwelling too much on this clarification of the intentions: the auditor or the management controller would, on the other hand, keep at it. For the evaluator, it is the content of the action that is more important than the initial objectives. However, their consideration can, to a certain extent, weigh on the choice of criteria by rejecting those that correspond to a goal unrelated to the programme's rationale. Thus the Community Policing³⁸ option implies the choice of a preventive or *ex ante* approach rather than a repressive or *ex post* one. Deterrence through surveillance will be favoured over waiting for the offense to be committed before intervening. The evaluator would otherwise be completely mistaken if he chose statistics pertaining to crime control as a measuring tool, recorded offences for instance: he would estimate the programme by the yardstick of a goal that it does not pursue. The difficulty is critical in the case of social prevention programmes that only indirectly aim at crime and insecurity reduction, through other objectives, such as stable job access programmes or the reduction of poverty. By evaluating them on the basis of objectives they do not pursue directly, the opportunity is missed of verifying whether they attain their immediate objectives or not. There remains the difficulty of determining whether the success of the immediate goal has ricocheted positively on the indirect objective, namely reduction of crime and insecurity.

Determining the resources is especially important for management auditors who want to measure the declared costs/outputs ratio from a managerial per-

³⁸ See, e.g., Skogan, 1990; Fielding, 1995; Rosenbaum, 1994; Skogan, Harnett, 1997a, b; Brodeur, 1998a, 2000.

spective³⁹. This is less important, it seems to me, for the evaluator. In any case, it involves identifying – not always an easy task – the new or additional resource specifically related to the programme. Often programmes function by redeploying existing funds or by tapping into constant or relatively constant resources. In such a situation, the additional cost is marginal, at least in its quantifiable aspect, for its implementation can demand extra effort, voluntarism and imagination on the part of the personnel, which is all the greater as the additional means given them are rather meagre⁴⁰.

In the evaluation, determining what has actually been implemented constitutes, it seems to me, the crucial point of the preliminary phase⁴¹. This is how the programme will not be blamed for failing because of defective or, simply, non-existent implementation. It is a well known fact that in this field, much more is said than done, and that announcements often are a substitute for action. To determine the nature of the implemented action without too much difficulty, it would be really useful to have settled on the evaluation even before the commencement of the programme; it is also vital that the evaluator be independent of both policy-makers and implementers; however, it is indispensable that he observe what is being done. This contradiction is not always easy to deal with: providing for the evaluation and its financing, having a central promoter select the projects to be assessed and approving their auditors are the very preventative measures a national decision-maker – legislative or executive – could take to oil the wheels.

Lastly, the preliminary phase of the evaluation ends with the estimation of the outputs. But this aspect is

³⁹ See, e.g. Braga et al., 1999.

⁴⁰ In this case, the launching of a new programme can bring in its wake – the evaluator should make note of this – a certain neglect of the customary tasks.

⁴¹ For example Brodeur, 1998b.

especially important for the programme management evaluator or for the auditor. By contrast, for the evaluator, delimiting the scope of the outputs specifically helps to distinguish them from the outcome, which constitutes the real criterion of evaluation⁴². Which is why the evaluator, contrary to the auditor or the programme management controller, cannot restrict himself to internal data; he needs a survey to constitute data that are external to the project and thus more apt to measure its outcome. Otherwise, according to Sherman (*et al.*, 1997), the effort is subjected to evaluation, not the results.

This is the stage where internal data is chiefly useful: to account for the implementation and its outputs. However the programme should not be trimmed down to a few indicators that would caricature it: only a comprehensive analysis can help to estimate the interventions' outcome⁴³. Moreover, a description of the operations is not enough either, they have to be replaced in their proper context before the replicability of the observed results can be estimated.

Hope (2009a) goes further and advises taking into account, not only the existence of an action, but also its intensity⁴⁴ by calculating an intensity score⁴⁵, borrowing the methodology from econometrics, an elegant way to measure the global substance of the action to be evaluated.

⁴² E.g. Crawford, 1998b; Hough, Tilley, 1998.

⁴³ See Pawson, Tilley, 1994.

⁴⁴ See also MacKenzie, 2006, 342.

⁴⁵ Time series analysis of programme impact.

VII - Evaluative know-how

A sort of minimal standard has been set comprising a before/after comparison, consideration of control groups and areas, and lastly, examination of the operation/impact relationship.

It is moreover on the basis of this yardstick that L. Van Noije and Karin Wittebrood (2008) have selected 69 valid evaluations out of the 149 they had managed to collect in the Netherlands. In so doing they explicitly refer to the precedents set by Sherman *et al.* (2002), *Campbell Collaboration Crime and Justice Group* (Farrington, Petrosino, 2001) and the *Scientific Model Scale* or SMS (Farrington, Petrosino, 2001 ; Farrington, 2003).

Founded and for many years presided by David Farrington, professor of Psychological Criminology at Cambridge University, the *Campbell Collaboration Crime and Justice Group*, proposes to undertake a systematic review of the evaluative studies conducted on the effectiveness of criminological interventions and make it widely accessible. It thus intends to continue Donald T. Campbell's (1969) *princeps* work, generally considered the promoter of evaluative research. We should note that these two major authors are both experimental psychologists: the quasi-experimental approach that they propose for evaluating crime prevention and safety policies owes much to the mode of reasoning and methods prevailing in their initial discipline.

Farrington (2003) retains five criteria for assessing the methodological quality of evaluations; he attributes the first four to Campbell:

- statistical conclusion validity refers to the critical nature of the relationship between the intervention and the presumed outcome;

- the internal validity refers to the examination of the chain of causation between the intervention and the outcome;
- the construct validity refers to the adequacy of the outcome with the operational definition and measurement of the theoretical constructs that underlie the intervention;
- external validity refers to the possibility of generalising the causal relationship between the intervention and the outcome;
- lastly, descriptive validity relates to the adequacy of the presentation, in a research report, of the key features of an evaluation.

Among all the scales pertaining to the methodological quality of evaluations, the most famous is the *Maryland Scientific Method Scale* (SMS) developed on the basis of the report prepared for the American Congress by Lawrence Sherman and his collaborators.

It constitutes five levels :

- level 1 : the simple correlation between a programme and the measurement of crime, at a single point in time;
- level 2 : temporal sequence between the programme and the crime outcome;
- level 3 : comparative measurement of crime between two or more comparable units of analysis, one with and one without the programme;
- level 4 : comparative measurement of crime between multiple units with and without the programme, controlling for other factors likely to influence crime;
- level 5 : random assignment and analysis of comparable units to programme and control groups.

But Farrington (2003) criticises it for not covering all the dimensions of validity retained by the *Campbell Group*.

The before/after comparison is obviously fundamental: without it there simply cannot be an evaluation. Four points at least deserve attention. First, evaluation should preferably be provided for before the commencement of the programme⁴⁶: the situation will be more easily observed *ex ante* rather than having to painfully reconstitute it *ex post*. Also, a sufficient number of criteria for this before/after comparison should be retained so as not to overlook unexpected effects. Such a precaution can be used for discerning the unintended effects more easily⁴⁷: harassing dealers can reduce the impact of drugs in an area, but police intervention methods exasperate the youth so much that violence increases. Spreading a before/after procedure beyond the sole implementation area of the programme allows for the identification of the possible displacement effects⁴⁸ of crime: it may decrease in areas where the programme has been implemented but can surface in an adjoining area; it also allows virtuous contagion effects⁴⁹ where investment in crime prevention is powerful enough to radiate in areas contiguous to its particular zone of intervention⁵⁰.

Specialists are highly critical of before/after measures that are not accompanied by the observation of control zones or groups where the programme to be assessed is not implemented. They even insist on the importance of having a pool of control zones or groups in order to neutralise the effect of a sudden crisis in one of them⁵¹. This amounts to an aspiration to move forwards from the rather primitive model of the 'black box'⁵² to a method

⁴⁶ E.g. Sherman et al., 1997; Brodeur, 1998b.

⁴⁷ E.g. Sherman et al., 1997; but also Hope (2009a) with regard to the Kirkholt programme.

⁴⁸ Hough, Tilley, 1998.

⁴⁹ Displacement effects are more often envisaged than expansion effects (Hope, 2009a).

⁵⁰ E.g. Ekblom et al., 1996.

⁵¹ See e.g. Bennet, 1991.

⁵² Said of an observation that considers the inputs and the outputs but not the intermediary mechanisms.

deemed quasi-experimental. For all that, the implementation of this procedure leads to massive problems. While it is possible to divide one neighbourhood or group into two and reserve one half for controlling, two difficulties emerge when one or several control areas or groups have to be located: i) it is difficult to find matching characteristics –two neighbourhoods seemingly similar in terms of their population and their problems, can nevertheless greatly differ in their histories and therefore their capacity for self-monitoring, and ii) another programme may be implemented in a control zone or population with effects similar to the effects resulting from the tested programme. However desirable it may be, comparability is often difficult to ensure. The aim of control zones or groups is to settle the issue: can the change observed be attributed to the programme or would it have occurred anyway despite its absence? But the functioning of this control is not automatic: thus the zero effect of a neighbourhood watch programme can simply be an expression of the similar levels of ‘social capital’ present in the experimental and control communities (Hope, 2009a). This type of difficulty obviously recedes with Tilley’s ‘realist’ evaluation: as it is no longer a matter of determining ‘what works’, but what has worked in a given context and in view of this context, it is obviously no longer necessary to find control zones or populations, nor expose oneself to the agony of this quest. From this perspective, it is not the programme that is being tested but the theory that underlies the actual implementation.

More generally, the mechanical application of a quasi-experimental procedure does not help to steer clear of all the selection biases that Hope (2009a) has specifically drawn our attention to. Thus, a community can be selected because it seems to favour the programme to be implemented, but the effect observed can be as much due to the ‘social capital’ present in this community, as to the programme in operation here⁵³. But specific de-

⁵³ Even though a programme can at least have a positive effect on the

prived zones or groups can also be selected and then what is observed is merely their reversion to the mean. Which is why Hope suggests modelling selection effects by drawing inspiration from micro-econometrics.

In any case, all these difficulties accord particular importance to the third phase of evaluation: there are three prerequisites to concluding as to the existence of an outcome, i) envisage and discard alternative explanations⁵⁴, ii) on the contrary, explain how the programmes actually implemented have arrived at the results observed⁵⁵ and iii) decide on the likelihood of this process. This is a crucial point where the evaluator's know-how and experience prove useful.

How about the downstream part of the evaluation, its utilisation? The monitoring of the programme should be considered separately from its replicability.

As a mere post-implementation judgement: the evaluation may only be of historical relevance to the programme, something which is not of particular interest to its directors and to the professionals. This is the real advantage of an impact study set up at the very outset: it is capable of providing information while underway, thus making it possible to adjust the tools⁵⁶.

mobilisation of the latent resources of the group or the area where it is being applied. It then becomes the intervening variable which produces the outcome, whereas the observation of a direct relationship between its implementation and the outcome is impossible.

⁵⁴ Rosenbaum, 1986; among other things, there has to be data over a sufficiently long period of the earlier situation to discern possible tendencies that accounted for the before/after changes, even if the programme had not existed (Hough, Tilley, 1998; Hope, 2000). The statistical significance of the quantitative changes observed has also to be checked (Sherman et al., 1997).

⁵⁵ See e.g. Tilley, 1993b; Hough, Tilley, 1998; Brodeur, 1998b.

⁵⁶ This is what the Chicago police reform programme evaluation aspired to do. See, Skogan, Hartnett, 1997a, b.

But it is also expected from the evaluation that it indicate solutions promising enough to be replicable. The huge report that Lawrence Sherman was commissioned to write for the American Congress (1997), primarily pursued this objective, as did also the report on correctional measures by MacKenzie (2006). It is in fact the reason why evaluation should focus on certain programmes, which by their rigour and scope are likely, in case of positive evaluation, to be re-used on a wider scale⁵⁷. Nevertheless, the exercise is a delicate one: to generalise a pilot-experience is not that straightforward; something that has given good results in a particular context can turn out to be less successful if transposed to other very different ones. Here, the selection biases can have a major impact, hence the importance of detecting and neutralising them before concluding as to the external validity of the programme under evaluation.

VIII - Conclusion

In the end, evaluation is a cumulative exercise. It is while evaluating that techniques may be improved. It is also while evaluating that practical knowledge of the impact of the different prevention measures is gradually mustered... knowledge that can always be revised, as is normally the case with scientific research, but knowledge that enables accounting for public policies and adjusting them as and when required. It is useless to try and evaluate everything, especially when experience and skills are modest. The outcome would be a pseudo-evaluation, based on impressions or trends and therefore doubtful. It is better to start modestly by selecting some specific programmes on which available resources and skills could be concentrated. Thus little by little, know-how will be developed and reasonably

⁵⁷ Hence, Hough, Tilley (1998) advise reserving it for innovative, elaborate and costly programmes; see also the proposal made by the Sherman report (et al., 1997).

reliable diagnoses accumulated. Thereafter the process will have a snowball effect through trial and error.

It should be reminded, however, that in the domain of evaluation the relationship between policy-makers and scholars is of a particularly delicate nature. Between the refusal of the former to approach the latter and, on the contrary, resort to a quasi-takeover, it is not easy to promote cooperation based on mutual respect of the independence of the two spheres. Without it however, evaluation can only be pretence.

For all that, the effectiveness of a programme does not settle the question of the relevance of its location: the resources that are expended in one place could be in short supply in another where the needs are more urgent. Over and above all evaluation, we cannot shirk the task of reflecting on how the target-areas of interventions are prioritised, otherwise security measures in the long run will merely be the privilege of the well-off⁵⁸.

Lastly, crime prevention and safety policies are on the whole incapable of preventing the devastating effects of an accumulation of negative socio-economic conditions on certain social groups or poverty-ridden urban zones⁵⁹. These policies should not help to mask the absence of effective social and economic policies or worse, the continued accumulation of segregative decisions and practices. Without an effective reaffiliation policy they would merely be an illusion.

This said, it would be interesting to find out how many (scientifically acceptable) evaluations have had a real change-inducing effect on public policies.

On the subject of recommendations, here are some sug-

⁵⁸ Crawford, 1998a, b; also Hope (2009a).

⁵⁹ Sherman et al. (1997) emphasise this fact repeatedly in relation to the United States.

gestions to anyone who would like to embark on the evaluation of crime prevention and safety policies:

- not to mix up evaluation – which applies to the impact of these policies on a target – with audit, programme controlling or cost-effectiveness calculations;
- to entrust the evaluation to a scientific body, competent and external to the institutions that are in charge of the programmes to be evaluated
- to respect the mutual exteriority of the policy-makers' and the evaluators' spheres;
- to plan the evaluation before the start of the programme;
- to provide data and know-how which are coherent with the nature of the evaluation.

Bibliography

BENNETT T., The effectiveness of a police-initiated fear-reducing strategy, *British Journal of Criminology*, 1991, 31, 1, 1-14.

BEAUVALLET M., *Les stratégies absurdes. Comment faire pire en croyant faire mieux ?*, Paris, Seuil, 2009.

BRAGA A.A., WEISBURD D.L., WARING E.J., GREEN MAZEROLLE L., SPELMAN W., GAJEWSKI F., Problem-oriented policing in violent crime places: A randomized controlled experiment, *Criminology*, 1999, 37, 3, 541-580.

BREEN E. (Dir.), *Évaluer la justice*, Paris, PUF, 2002.

BRODEUR J.-P. (Ed.), *How to Recognize Good Policing: Problems and Issues*, Thousand Oaks, Sage, 1998a.

BRODEUR J.-P., L'évaluation des programmes d'intervention sécuritaire ; quelques remarques sur les expériences canadienne et américaine, in FERRET F., OCQUETEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998b, 13-22.

BRODEUR J.-P., *Les réformes de la police en Amérique du Nord : bilan des évaluations*, Eines para l'analisi de la seguretat ; Estat de la recerca, Mollet del Vallès, Escola de policia de Catalunya, 2000.

CAMPBELL D.T., Reforms as Experiments, *American Psychologist*, 1969, 24, 4, 409-429.

COOK T.D., CAMPBELL D.T., *Quasi-Experimentation*, Chicago, Rand-Mc Nelly, 1979.

CRAWFORD A., *Crime Prevention and Community Safety. Politics, Policies and Practices*, Longman, 1998a.

CRAWFORD A., Culture managériale de l'évaluation et responsabilités ; quelques leçons de l'expérience britannique des programmes locaux de sécurité in FERRET F., OCQUETEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998b, 51-82.

DONZELOT J., ESTEBE PH., *L'État animateur. Essai sur la politique de la ville*, Paris, Éd. Esprit, 1993.

EKBLOM P., Safer cities and residential burglary: A summary of evaluation results, *European Journal of Criminal Policy and research*, 1996, 4, 1, 22-52.

EKBLOM P., Évaluation du programme *Safer Cities* ; impact et processus, in FERRET F., OCQUETEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998, 27-50.

EKBLOM P., LAW H., SUTTON M., CRISP P., WIGGINS R., *Safer Cities and Domestic Burglary*, Home office Research Study 164, London, Home Office research and Statistics Directorate, 1996.

FARRINGTON D.P., Methodological Quality Standards for Evaluation Research, *Annals of the American Academy of Political and Social Science*, 2003, 587, 49-68.

FARRINGTON D.P., PETROSINO A., The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science*, 2001, 587, 35-49.

FIELDING N.G., *Community Policing*, Oxford, Clarendon, 1995.

HOPE T., *Counting on the Counter-Factual? Evaluation and Crime Reduction in Risk Society*, Leicester, British Conference of Criminology Conference, 2000.

HOPE T., *Evaluation of Safety and Crime Prevention Policies. England and Wales*, FP6, CRIMPREV, WP7, WS3, 2009a [www.crimprev.eu].

HOPE T., *The evaluation of policies from a methodological perspective - Discussion*, FP6, CRIMPREV, WP4, WS6, Universidade de Porto, Faculty of Law, Escola de Criminologia, 2009b.

HOUGH, M., MAYHEW P., L'évolution de la criminalité à travers deux décennies du British Crime Survey, *Dévi-ance et Société*, 2004, 28, 3, 267-284.

HOUGH M., TILLEY N., *Auditing Crime and Disorder: Guidance for Local Partnerships*, London, Home Office, 1998.

JUNGER-TAS J., *Meta-evaluation of Dutch Crime Prevention programs*, Den Haag, Ministry of Justice, WODC, 1989.

KESSLER M.C., LASCOUMES P., SETBON M., THOENIG J.-C. (dir.), *L'évaluation des politiques publiques*, Paris, L'Harmattan, 1998.

LAGRANGE H., SINA F., Réflexions de méthode sur les diagnostics locaux de sécurité, in FERRET F., OCQUE-TEAU F. (Eds), *Évaluer la police de proximité ? Problèmes, concepts, méthodes*, Paris, La Documentation française, 1998, 91-114.

LASCOUMES P., SETBON M., *L'évaluation pluraliste des politiques publiques ; enjeux, pratiques, produits*, Paris, CGP, 1995.

LEIGH A., *Policing in England and Wales: National Developments in Performance Management*, Congrès internacional sobre les eines per a l'anàlisi de la sug-eretat. L'estat de la recerca, Mollet del Vallès, Escola de policia de Catalunya, 2000.

MACKENZIE D.L., *What Works in Corrections. Reducing the Criminal Activities of Offenders and Delinquents*, Cambridge, Cambridge University Press, 2006.

MARTINSON R., What works? Questions and answers about prison reform, *Public Interest*, 1974, 35, 22-45.

PATTON M.Q., *Qualitative Evaluation and Research Methods*, Newbury Park, Ca, 1980.

PAWSON R., TILLEY N., What works in evaluation research?, *British Journal of Criminology*, 1994, 34, 3, 291-306.

PAWSON R., TILLEY N., *Realistic Evaluation*, London, Sage, 1997.

POLDER W., VAN VLAARDINGEN F.J.C., Preventiestrategieën in de praktijk, een meta-evaluatie van criminaliteitpreventieprojecten, *Justitiele Verkenningen*, 1992, 2, 5-104.

ROBERT PH., Évaluer la prévention, *Archives de politique criminelle*, 1994, 16, 53-70.

ROBERT PH., The evaluation of prevention policies, *European Journal of Crime, Criminal Law and Criminal Justice*, 2003, 1, 114-130.

ROBERT PH., *L'évaluation des politiques de prévention et de sécurité examinée sous une perspective méthodologique*, FP6, CRIMPREV, WP4, WS6, Universidade de Porto, Faculty of Law, Escola de Criminologia, 2009.

ROBERT PH., LASCOUMES P., *Les bandes d'adolescents ; une théorie de la ségrégation*, Paris, Éditions ouvrières, 1974.

ROCHÉ S., *Police de proximité. Nos politiques de sécurité*, Paris, Seuil, 2005.

ROSENBAUM D.P. (Ed.), *Community Crime Prevention; Does It Work?*, Beverley Hills, Sage, 1986.

ROSENBAUM D.P., *The Challenge of Community Policing: Testing the Promises*, Thousand Oaks, Sage, 1994.

SHERMAN L.W., FARRINGTON D.P., WELSH B.C., MACKENZIE D.L. (Eds), *Evidence-based crime prevention*. London/New York, Routledge, 2002.

SHERMAN L.W., FARRINGTON D.P., WELSH B.C., MACKENZIE D.L. (Eds), *Evidence-based crime prevention (revised edition)*, London, Routledge, 2006.

SHERMAN L.W., GOTTFREDSON D., MAC KENZIE D., ECK J., REUTER P., BUSHWAY S., *Preventing Crime: What Works, What Doesn't, What's Promising*, Washington DC, NIJ, 1997.

SKOGAN W.G., *Disorder and Decline: Crime and the Spiral of Decay in American Cities*, NY, Free Press, 1990.

SKOGAN W.G., HARTNETT S.M., *Community Policing, Chicago Style*, Oxford, Oxford University Press, 1997a.

SKOGAN W.G., HARTNETT S.M. AND THE CHICAGO COMMUNITY POLICING EVALUATION CONSORTIUM, *Community Policing in Chicago, Year Four: An Interim Report*, Illinois Criminal justice Information Authority, 1997b.

SMEETS S., TANGE C., L'évaluation des politiques de sécurité et de prévention en Belgique, FP6, CRIMPREV, WP7, WS3, 2009 [w ww.crimprev.eu].

SOETENHORST DE SAVORNIN LOHMAN J., New Forms of crime Prevention: the Dutch Experience, in HEBBERECHT P., SACK F. (dir.), *La prévention de la délinquance en Europe ; nouvelles stratégies*, Paris, L'Harmattan, 1997, 83-100.

TILLEY N., *Understanding Car Parks, Crime and CCTV: Evaluations and Lessons from Safer Cities*, London, HMSO, 1993a.

TILLEY N., *After Kirkholt – Theory, Method and Results of Replication Evaluations*, Crime prevention Unit Series, Paper 47, London, Home Office, 1993b.

TILLEY N., *Thinking about Crime Prevention Performance Indicators*, London, Home Office, 1995.

VAN DIJK J.J.M., The State of Knowledge in the Netherlands, in ROBERT PH. (Ed.), *Crime and Prevention Policy*, Eigenverlag Max-Planck-Institut für ausländisches und internationales Strafrecht, 1993, 243-263.

VAN NOIJE L., WITTEBROOD K., *Sociale veiligheid ontseuteld: veronderstelde en werkelijke effecten van veiligheidsbeleid*, Den Haag, Sociaal en Cultureel Planbureau, 2008.

WITTEBROOD K., *Evaluation of safety and crime prevention policies in the Netherlands*, FP6, CRIMPREV, WP7, WS3, 2009 [www.crimprev.eu].

WYVEKENS A., Quand l'urbain rencontre la sécurité : la prévention situationnelle en France aujourd'hui, *Revue de droit pénal et de criminologie*, 2008, 9-10, 897-900.

WYVEKENS A., *L'évaluation des politiques de prévention et de sécurité en France : une activité problématique*, FP6, CRIMPREV, WP7, WS3, 2009 [www.crimprev.eu].

ZAUBERMAN R., ROBERT PH., POTTIER M.-L., Risque de proximité ou risque lié au style de vie ; enquêtes et évaluation de la sécurité urbaine, *Cahiers de la sécurité intérieure*, 2000, 42, 193-220.