

# L'évaluation de l'action policière en Amérique du Nord

Par  
**Jean-Paul Brodeur**  
Professeur/chercheur  
Centre international de criminologie comparée (CICC)  
Université de Montréal

**Juin 2002**

## *0. Introduction*

Depuis 1990, au moins, l'évaluation de la prestation de service de la police a été conduite de manière systématique. Il y a trois raisons qui expliquent le développement de la recherche évaluative :

a) La première raison tient à la réforme des opérations policières dans le monde anglo-saxon (cette réforme s'est, bien sûr, maintenant étendue à plusieurs pays). Elle a donné lieu à l'élaboration d'une méthode d'intervention qu'on a dénommée la méthode SARA. Chacune des lettres de cet acronyme se réfère à une opération, soit S pour description de la SITUATION, A pour ANALYSE des problèmes, R pour détermination de la RÉPONSE et A pour l'APPRÉCIATION, ou encore l'évaluation des résultats de l'intervention policière. On le voit, l'évaluation des résultats d'un programme d'intervention policière est une partie intégrante de celui-ci.

b) Dans les pays anglo-saxons et, de façon plus particulière aux États-Unis, des organismes gouvernementaux attribuent d'importantes subventions à divers corps policiers pour améliorer les services qu'ils fournissent à la population. Ces organismes sont soucieux de subventionner des programmes d'intervention qui ont du succès. C'est une autre des raisons pour lesquelles les études évaluatives se sont développées dans les pays anglo-saxons. En 1996, le professeur Lawrence Sherman a pris la tête d'une équipe de l'Université du Maryland, qui a fait pour le congrès des États-Unis le bilan de tous les programmes d'intervention en matière de justice pénale qui avaient été préalablement évalués. Cette équipe a recensé 101 évaluations des prestations de la police américaine, qui répondaient aux critères de rigueur méthodologique retenus. Au Royaume-Uni, on compte plus d'une

quarantaine de ces évaluations, plusieurs étant subventionnées par le Home Office Research Unit.

c) la troisième raison tient au fait que ce ne sont pas seulement les interventions de la police publique qui sont évaluées, mais aussi celles des agences de sécurité privée. Plusieurs de ces évaluations ont été publiées dans des revues telles que *The Journal of Security Administration* et *The Security Journal*. On comprend immédiatement l'intérêt du secteur privé pour une évaluation du succès de ses opérations.

Nous ne nous proposons pas ici de vous présenter le résultat de ces évaluations, l'ayant déjà fait ailleurs (Brodeur, sous presse). Nous discuterons plutôt des problèmes que soulève en elle-même l'évaluation de la prestation de service de la police. Ces problèmes sont nombreux et de divers ordres. Notre exposé est divisé en trois parties : nous présenterons d'abord des remarques sur l'objet et sur les méthodes d'évaluation. En second lieu, nous ferons état d'un certain nombre de difficultés rencontrées par l'évaluation. Dans la dernière partie, la plus longue, nous vous soumettrons un ensemble de suggestions tant méthodologiques que pratiques pour résoudre les problèmes identifiés auparavant.

## **1. Objet et méthode de l'évaluation**

Lorsque nous nous penchons sur l'objet de l'évaluation, il convient d'abord de distinguer entre l'intervention de sécurité et l'intervention de police. L'intervention de sécurité n'est pas réservée aux services publics, la part du privé y étant de plus en plus considérable (Johnston, 1992). Surtout, l'intervention de sécurité est d'une double nature : elle peut être effectuée par des ressources humaines (des personnes) ou par le moyen de ressources technologiques, comme par exemple un détecteur de métal dans les aéroports. L'intervention de police, tel que nous entendons ces mots, est pour l'essentiel effectuée par la police publique, souvent en partenariat avec divers éléments de la société civile. Dans la suite de cet exposé, nous comptons donc nous référer avant tout à l'intervention de police. Voici en peu de mots pourquoi : les évaluations qui ont été conduites des interventions de police sont plus fiables que celles qui ont porté sur les interventions de sécurité, qui ne sont pas toujours au-dessus du soupçon de gonfler la qualité des services ou de mousser la vente d'un produit technologique.

### **1.1 Objet**

Un examen des recherches effectuées sur l'intervention de police révèle une première chose. Certaines de ces recherches portent sur une intervention de grande envergure (en fait, sur une masse d'interventions diverses). Par exemple, l'étude de Skogan et Hartnett (1997) s'est déroulée pendant plus de deux ans et ses auteurs ont procédé à plus de 7 000 entrevues avec des policiers de Chicago. Il s'agissait dans ce cas de l'application d'un nouveau *style* d'intervention, qui recouvrait la plupart des dimensions de l'action policière. D'autres recherches, et ce sont les plus fréquentes, portent sur un programme d'intervention défini de façon relativement étroite dans son envergure et dans son application dans le temps. Toutefois, que ce soit tout un style d'intervention qui est évalué ou seulement un programme, autant le style sous examen que les divers programmes participent d'une même perspective sur la façon de faire la police. Il s'agit de cette réforme qui a été définie par Herman Goldstein (1990) sous l'appellation de « Police de résolution de problèmes » (PRP), qui a été longtemps associée avec la police de communauté (*community policing*, connue en France sous l'appellation de police de proximité). Ce style a été redéfini au Royaume-Uni comme « *Intelligence-led Policing* » (ILP; en français, « Police alimentée par le renseignement »). Le renseignement s'entend ici en son sens le plus large comme information sur les aspects du problème à résoudre. Quelles que soient ces variantes – PRP, ILP ou même *community policing* – l'évaluation des activités conduites dans cette perspective porte sur un objet : la qualité de l'intervention de police. On le subdivise toutefois en quatre composantes.

a) **La validité de la définition du problème et la qualité de l'expertise** à partir de laquelle cette définition du problème a été élaborée. L'une des maximes de la PRP est que tout est déjà largement joué au stade de la définition du problème. Nous vous fournirons par la suite quelques exemples au soutien de cette proposition. En effet, la détermination des modalités de l'intervention de police dépend essentiellement de la caractérisation de la nature du problème. Une des questions que l'on peut se poser à ce stade est, par exemple, celle de savoir si la population qui fait l'objet de l'intervention a été consultée sur sa vision des problèmes du territoire.

b) **La perception par le public de l'intervention de police** : nous ne voulons pas entendre par ce terme de perception le *judgement* qui est porté par le public sur l'intervention de police (on pourra toujours mesurer ce jugement plus tard). Nous entendons par « perception » la simple réalisation par le public que les policiers font les choses de façon différente. En d'autres termes, le public s'est-il même aperçu du changement dans les modes d'intervention ? On connaît des cas notoires où le public ne s'était même pas aperçu que les policiers tentaient d'implanter un nouveau programme d'intervention. Le problème de fond que soulève ce manque d'aperception est qu'il ne retient pas les gens de juger d'un programme. En effet, ceux-ci sont gênés de reconnaître qu'ils ne se sont pas aperçus qu'on mettait un nouveau programme d'intervention en place. De façon générale, les personnes sondées tentent de compenser leur ignorance d'un programme en lui attribuant une note moyenne ou une bonne note.

c) **L'application par les intervenants du programme d'action** : il s'agit encore une fois d'une étape qui est souvent négligée. On tente de mesurer les résultats d'une intervention sans s'être interrogé sur la qualité de sa réalisation. La somme des facteurs affectant l'évolution d'une conjoncture relative à la sécurité est en effet si capricieuse qu'on ne saurait écarter que cette conjoncture évalue favorablement, quand bien même l'intervention policière prévue se serait mal passée ou qu'elle n'aurait pas effectuée du tout. À cet égard, les recherches évaluatives ont identifié un effet de diffusion dans l'espace et dans le temps des bénéfices d'un programme : un quartier qui n'a pas fait l'objet d'une intervention de police peut en partager de façon malheureusement toute provisoire les bénéfices par effet de diffusion. Il est donc impératif de vérifier dans quelle mesure un programme d'intervention a été appliqué en conformité avec sa définition par les intervenants sur le terrain avant de procéder à l'estimation de son impact.

d) **L'impact ou encore les effets de l'action entreprise** : nous n'en dirons pas plus long sur cette dernière dimension de l'évaluation qu'on peut intuitivement comprendre sans peine. Cette mesure est toutefois la plus complexe de toutes et nous commencerons d'en discuter en disant quelques mots des problèmes de méthode.

## 1.2 Méthode

Le premier principe général de méthode est que l'évaluation doit être effectuée par une instance *autre* que celle qui commande le programme soumis à l'évaluation et *indépendante* de celle-ci. L'expérience a montré que les procédures d'auto-évaluation péchaient par trop grande complaisance (Bratton, 1998; Silverman, 1999).

Si l'évaluation est menée par une partie extérieure à l'intervention de police, elle doit néanmoins esquiver une embûche, surtout lorsque cette évaluation est effectuée en même temps que l'intervention de police, comme cela est souhaitable. À mesure que les résultats de l'évaluation s'accumulent, ils doivent être dans la mesure du possible communiqués aux intervenants. Ceci permet à ces derniers de rectifier au besoin leur pratique. De façon plus importante, peut-être, la communication des résultats au fur et à mesure qu'ils sont produits évite qu'un fossé s'établisse entre les évaluateurs et les intervenants. *Le second principe général de l'évaluation serait donc d'éviter que les évaluateurs apparaissent aux intervenants comme un corps d'inspecteurs distribuant des sanctions* : il importe qu'ils soient perçus comme une importante ressource de la pratique.

On peut finalement énoncer un troisième principe général de l'évaluation : *plus une intervention est conduite en partenariat et plus elle est difficile à évaluer*. Lorsqu'on trouve plus de deux partenaires, il devient à toutes fins utiles presque impossible d'évaluer la contribution respective des divers partenaires.

Il reste enfin un ensemble de principes méthodologiques de l'évaluation qui sont plus techniques et qui tiennent en des règles de procédure entendues dans un sens épistémologique plus étroit. L'idéal serait ici que l'évaluation tentât de reproduire les conditions d'une expérience scientifique et qu'elle fût, en d'autres mots, articulée selon une conception expérimentale. En réalité, il est rarissime que cet idéal expérimental puisse être atteint. Nous distinguerons donc, en accord avec l'équipe réunie par le professeur Sherman, cinq niveaux de rigueur, en commençant par le plus minimal.

- Il arrive qu'en conduisant une évaluation, on soit requis d'intervenir *après* le début d'un programme d'intervention, alors qu'il semblerait de rigueur tout à fait élémentaire que l'on puisse comparer une situation *avant* le déroulement d'une intervention et *après* que celle-ci se soit déroulée. Aussi élémentaire qu'elle soit, cette exigence ne peut pas toujours être respectée, les intervenants et leurs commanditaires n'étant pas toujours sensibles aux exigences de la méthode. Voici donc ce qui tient lieu de substitut à cette exigence élémentaire : les effets observés à la fin d'une intervention, dont on n'a pu observer les conditions initiales, doivent absolument se retrouver dans *plusieurs sites ayant fait l'objet d'interventions analogues*. À défaut d'une procédure distinguant l'avant et l'après, on doit donc observer les effets d'un programme en plusieurs endroits pour tirer quelque conclusion que ce soit sur la validité de celui-ci. En clair, on ne peut tirer aucune conclusion des résultats constatés en milieu ou en fin d'intervention sur un seul site.

- La condition méthodologique la plus généralement revendiquée pour conduire une évaluation est que l'on effectue une description de la situation *avant* l'application du programme, pour pouvoir mesurer avec un début de rigueur l'impact effectif du programme *après* son application. La progression dans la rigueur consistera par la suite à ajouter des exigences encore plus strictes de méthode à *la conception de base de type avant / après*.

- Le palier ultérieur dans la rigueur consiste en effet à ajouter ce qu'on appelle un *groupe contrôle*. L'évaluation procède alors de la façon suivante : on commence par décrire les conditions qui prévalent, par exemple dans deux quartiers comparables, et on applique un programme d'intervention dans l'un d'eux seulement. On mesure ensuite si la situation finale est différente dans le quartier qui a fait l'objet de l'intervention en comparant celle-ci à celle qui prévaut dans le quartier contrôle. *Ce concept de recherche qui repose sur des mesures avant / après, avec en plus un groupe contrôle, est celui qu'on doit s'efforcer d'appliquer*. La plupart des recherches conduites sur l'intervention policière et même sur l'intervention de sécurité dépassent rarement ce niveau.

- Nous ne mentionnons que par souci d'exhaustivité les deux paliers supérieurs de la rigueur méthodologique: (1) On peut conduire le type de recherche précédemment décrit sur plusieurs sites (plusieurs quartiers d'intervention appariés respectivement à leur quartier contrôle); (2) finalement, on peut s'assurer que les sujets qui font l'objet de l'intervention sont choisis rigoureusement au hasard; de façon analogue, les sujets sondés sur la qualité du programme d'interventions sont aussi sélectionnés au hasard. Avec cette procédure d'allocation au hasard, on se hisse au palier de la recherche expérimentale au sens strict du terme. Très peu d'études sur la police et la sécurité parviennent à ce niveau.

Il est un dernier problème dont nous traiterons dans la partie suivante : il s'agit du choix des indicateurs de l'effet du programme. L'indicateur le plus souvent utilisé est la statistique des crimes et des délits rapportés à la police, dont on décrit les fluctuations en fonction du programme d'intervention. On ne doit certes pas s'abstenir d'utiliser la statistique criminelle officielle mais on ne saurait se reposer sur elle de façon exclusive.

## **2. Les difficultés rencontrées dans la pratique de la recherche évaluative**

Nous avons commencé par dire que la composante « évaluation » était intégrée à un style de police – la PRP – comme la dernière étape d'un processus. Il s'en faut toutefois de beaucoup pour que cette intégration se soit effectuée dans tous les cas avec satisfaction. En réalité, lors d'entretiens avec Herman Goldstein, celui-ci nous a confié que la composante « évaluation » était celle qu'il avait le moins développée jusqu'ici dans ses travaux. Un certain nombre de difficultés se sont donc révélées à l'usage.

## **2.1 Organisation, processus ou personnes**

Nous nous sommes déjà penché sur l'objet de l'évaluation en discutant de son étendue plus ou moins grande. Il reste une difficulté qui n'a pas jusqu'ici été abordée : le terme « style d'intervention », que nous avons utilisé pour désigner l'objet de l'évaluation est en réalité ambigu : il peut désigner une organisation policière, un processus d'intervention plus ou moins large ou la prestation de diverses *personnes*. Sans rien dire des difficiles problèmes qui sont soulevés par la politique des organisations (une organisation n'apprécie pas être estimée mal performante), il faut insister sur la redoutable ambivalence selon laquelle l'évaluation peut porter sur des entités relativement impersonnelles, comme des systèmes, des processus ou des organisations, et également porter sur des *personnes*. Dans ce dernier cas, on comprendra que celles-ci soient réticentes à participer à une évaluation dont elles pourraient éventuellement faire les frais, au plan de leur carrière. Il existe diverses stratégies pour atténuer ce problème, mais aucune d'elles ne parvient à le résoudre complètement.

## **2.2 Les indicateurs de l'évaluation**

Il faut le reconnaître sans ambages, ce sont les statistiques officielles des crimes rapportés à la police qui constituent le plus souvent le principal indicateur de la prestation de service de la police. La médiatisation de la statistique officielle a rendu son utilisation *incontournable* pour les fins de l'évaluation. L'utilisation de cette mesure n'en comporte pas moins un très grand nombre de difficultés, dont nous n'avons pas le temps de traiter (pour une analyse approfondie, voir Ocqueteau, 2000). Soulignons toutefois le principal paradoxe de cet indicateur : plus un corps policier suscite la confiance de la population et plus cette dernière aura tendance à lui rapporter la perpétration d'infractions, de telle sorte qu'une augmentation des chiffres de la statistique des crimes rapportés à la police peut désigner un accroissement de la qualité du service policier, et non pas une diminution de celle-ci.

Dans les grandes évaluations qui ont été effectuées aux États-Unis (par exemple, dans les villes de Chicago et de Jersey City), on a utilisé quatre types d'indicateurs : (1) des entrevues / sondages auprès de la population; (2) des entrevues / sondages auprès des policiers et, le cas échéant, de leurs partenaires; (3) les statistiques officielles des crimes rapportés à la police; (4) des sondages locaux de victimation, dont les résultats sont comparés aux statistiques officielles. Il est relativement rare qu'on utilise toutes ces mesures en même temps, le coût de leur utilisation conjuguée étant trop élevé. Mentionnons cependant que lorsqu'ils rencontrent des difficultés à instaurer le changement, certains corps policiers ont tendance à recourir de façon fréquente aux entrevues / sondages auprès de leur personnel. Ce fut ainsi le cas à Montréal de 1990 jusqu'à maintenant. Cette tendance s'avère d'autant plus qu'une force policière est syndiquée et que les syndicats sont militants (le service et le syndicat rivalisent alors de sondages du personnel).

## **2.3 Problème particulier de méthode**

Après avoir traité de façon plus générale de la méthode dans la première partie de cet exposé, nous aimerions signaler trois problèmes plus spécifiques

- **La terminologie juridique** : Il n'est pas souhaitable de s'en remettre essentiellement aux catégories juridiques pour définir un problème, car celles-ci sont trop polyvalentes. Par

exemple, un incendie criminel peut être une manifestation de la frustration des jeunes; ce peut être un crime dont la motivation est religieuse; ce peut être un règlement de comptes entre gens du milieu; ce peut être également la destruction d'une scène de crime où s'est produit un autre crime grave (par exemple, un meurtre) et ce peut être aussi une fraude perpétrée contre des assureurs. Cette énumération n'épuise pas la signification de l'incendie criminel. La polyvalence de cette catégorie se retrouve dans de nombreux autres types de crimes.

- **Corrélation et causalité** : Ce problème est sans doute le plus classique de l'évaluation. De façon générale, la mesure d'un impact s'efforce d'établir une concomitance entre des interventions policières et un fléchissement dans la courbe des crimes faisant l'objet de cette intervention. Comme l'ont remarqué Pawson et Tilley (1994), la concomitance entre deux séries temporelles ne nous indique pas de façon précise quel est l'aspect opérant d'une intervention, qui produit le résultat constaté. Pour le dire autrement, *lorsqu'on en reste à l'établissement de corrélations, la courbe statistique de la criminalité et la ligne de l'intervention policière deviennent des parallèles dont on ne sait comment elles se rejoignent pour évoluer de concert.*

- **Implantation et impact** : Une évaluation détaillée de la mise en application d'un programme de PRP à New York (McElroy, 1993) a révélé une difficulté imprévue. Pour l'essentiel, la mise en application du programme d'intervention policière a eu pour effet de redéfinir les problèmes qu'on voulait résoudre à l'origine, les policiers étant les parties prenantes dans cette redéfinition. Il s'est donc révélé impossible de mesurer l'impact des interventions sur les problèmes, tels qu'ils avaient été initialement définis. Cette aporie est difficile à résoudre : de façon ultime, elle signifie que les premiers stades de la mise en application d'un programme d'intervention modifient les paramètres de la mesure de son impact, qui constitue le dernier terme de l'évaluation et sa principale raison d'être. L'évaluation se trouve ainsi toujours reportée vers son commencement.

## 2.4 Les coûts

Cette difficulté est sans doute la plus banale mais elle est également la plus décisive. De 1997 à 1999, le *National Institute of Justice* (NIJ) des États-Unis a subventionné treize projets d'évaluation de programmes de PRP ou de *community policing*. La plus élaborée de ces évaluations a coûté 576 000 (*US dollars*) et la moins développée 250 000 \$ : le coût moyen d'une évaluation fut de 324 279 \$. Ces coûts sont exorbitants et s'expliquent en partie par la volonté de la NIJ de faire la preuve que la PRP et la police de communauté étaient des stratégies viables; ils s'expliquent aussi par la compétition à laquelle se sont livrées les diverses villes américaines pour paraître parmi les plus performantes.

## 3. Le savoir et l'expérience

À l'exception des États-Unis, il est improbable que l'on puisse consacrer à l'évaluation des prestations de la police des sommes d'argent aussi importantes que celles que nous avons mentionnées plus haut. En réalité, nos propres études sur l'évaluation nous ont révélé que quels que soient les argents consacrés à ces évaluations, il y en avait très peu qui respectaient les exigences d'une conception expérimentale. C'est pourquoi nous nous proposons dans cette dernière partie de notre exposé de distinguer entre la constitution d'un *savoir* accumulé sur l'évaluation et entre la poursuites d'expériences quasi scientifiques proprement dites. Quelles que soient nos difficultés à réaliser une expérience d'évaluation qui soit effectuée en respectant toutes les règles de l'art, il est peut-être possible d'extraire d'un examen attentif d'un grand nombre de recherches d'évaluation, dont aucune ne réaliserait en elle-même l'idéal expérimental, un ensemble de connaissances pertinentes qui pourraient se révéler *en pratique* d'une grande utilité pour l'évaluation raisonnée des prestations de la police. En d'autres termes, nous faisons le pari qu'un ensemble de tactiques (d'expériences) imparfaites puisse être structuré de manière à constituer une stratégie d'évaluation pragmatiquement satisfaisante.

Voici donc un ensemble de procédés peu coûteux que l'on peut mettre à contribution pour se constituer une stratégie d'évaluation.

### **3.1 Les leçons indirectes**

On peut extraire de recherches évaluatives méthodologiquement insatisfaisantes un ensemble de *notions* relativement indépendantes du contexte où elles ont été initialement élaborées et qui peuvent trouver une application pratique ailleurs. La condition d'emploi de ces notions est qu'on puisse en observer les manifestations de façon récurrente dans un ensemble de recherches qui deviennent significatives davantage par leur convergence que par leur qualité intrinsèque. Nous citerons deux exemples de ces notions.

Le premier est bien connu : il s'agit de la notion de déplacement ou de transfert de la criminalité. On constate une diminution considérable d'une forme de délinquance dans un quartier, pour s'apercevoir ensuite que la criminalité n'a fait que se déplacer dans un quartier adjacent. De façon analogue, les recherches en prévention situationnelle (Clarke, 1992) ont constaté l'existence d'un phénomène de diffusion des bénéfiques, qui est le pendant positif du déplacement. Par exemple, un programme d'intervention peut être initialement élaboré pour cibler le vol à la roulotte dans un quartier et donner par la suite d'autres dividendes par rapport au cambriolage ou au vol à l'étalage.

### **3.2 Les narrations suggestives**

Le prototype de ce que nous appelons une narration suggestive a été produit par Clifford Shearing et Philip Stenning (1985) dans leur description d'un programme de contrôle des mouvements de foule complètement intégré à l'environnement. L'environnement dont il est question est celui du parc d'amusement Disneyland, à Orlando, en Floride. Le travail de ces deux auteurs n'a d'aucune façon consisté à évaluer l'impact des mesures de sécurité utilisées à Disneyland. Il a acquis cependant le statut d'un classique parce qu'il nous montre une façon radicalement alternative d'exercer un contrôle sur le déplacement des foules : celui-ci ne s'effectue pas au moyen d'un ensemble d'agents de police ou de sécurité mais au moyen de l'architecture physique d'un environnement qui structure selon un programme antécédent les déplacements d'une foule. Il faut à cet égard souligner qu'une impulsion importante a été donnée à la prévention du crime par les travaux en architecture d'Oscar Newman (1972). À partir de ses recherches sur des projets de logements sociaux réalisés aux États-Unis, Newman a montré que des environnements tels que les « cités-dortoirs » maximisaient les risques de délinquance.

### **3.3 Les effets physiquement perceptibles**

On doit faire une distinction entre des changements qui sont perceptibles en quelque sorte à l'œil nu, et des effets moins manifestes qui ne peuvent être saisis que par la statistique. Par exemple, le cambriolage n'est pas un crime qui se montre à d'autres qu'à sa victime. En conséquence, sa croissance ou sa décroissance ne sont respectivement perceptibles qu'au moyen de la statistique de la délinquance rapportée ou au moyen de sondages de victimation.

Il n'en va pas ainsi de tout un ensemble de délits dont le profil de visibilité est très élevé : l'émeute urbaine, les incendies de voitures, le vandalisme ou une certaine délinquance de mœurs qui se produit dans des lieux publics (prostitution). On se réfère aux États-Unis et au Canada à cette délinquance de visibilité sous le nom de désordre (physique, social ou sexuel; Skogan, 1990). Dans le cas de ce dernier type de délinquance, il se peut que l'évaluation du succès d'un programme puisse s'effectuer par le moyen de recherches de terrain à méthodologie rudimentaire : on constate en première part la disparition ou la décroissance du phénomène indésirable par suite d'un programme d'intervention et on s'assure qu'il ne s'est pas simplement déplacé dans un quartier avoisinant.

Les exigences les plus élevées, pour ce qui est de la recherche évaluative, ont été formulées à l'occasion de recherches portant sur l'impact d'une mesure sur une délinquance

en retrait de l'espace public et dans cette mesure peu visible ou invisible (par exemple, la violence conjugale ou la délinquance sexuelle familiale). Dans ce cas, toutes les précautions méthodologiques sont nécessaires et elles ne sont pas toujours suffisantes. On a toutefois tort de penser qu'il doit en être ainsi dans tous les cas de programmes s'attaquant à des comportements nocifs qui s'affichent. Dans ce dernier cas, les progrès peuvent faire l'objet d'un constat empirique relativement brut.

### 3.4 Les écarts statistiques considérables

Certains effets ne se révèlent toutefois qu'à la statistique. Il peut arriver que la mise en place d'un programme d'intervention ait pour résultat des écarts statistiques *très considérables* par rapport à une situation antérieure, même si celle-ci n'a pas été décrite avec toute la rigueur méthodologique souhaitable. Par exemple : les statistiques sur les vols à la roulotte dans une zone circonscrite pour les fins d'un programme d'intervention peuvent par hypothèse décroître du double au triple à mesure que le programme est mis en place. Si cette diminution se maintient, et si l'on ne constate pas d'effet de déplacement, on pourrait alors valablement conclure que l'intervention a du succès même si l'on n'a pas effectué cette vérification au moyen d'une recherche de nature expérimentale. De tels écarts mentent rarement de façon assidue.

En d'autres termes, dans le cas des écarts statistiques sensibles, comme dans celui de la décroissance de la délinquance affichée, on trouve des phénomènes de surface dont l'existence peut être constatée sans que l'on déploie un appareil méthodologique lourd. Il faut toutefois compenser ce qu'on épargne du côté de l'appareillage méthodologique par un sens critique aigu. Ainsi, pour ce qui est de l'exemple cité plus haut, il faut s'assurer que les écarts statistiques considérables qui sont constatés ne soient pas le produit d'une fabrication. Nous rajouterons à cet égard que les manipulations statistiques dont on soupçonne les agents du contrôle social sont moins fréquentes et d'ampleur moins considérable que ce qu'on imagine. Les manipulations les plus grossières se trouvent plutôt du côté du « packaging » (présentation graphique) de l'information (Tufte, 1983) et de l'interprétation des variations de la statistique des crimes rapportés à la police.

### 3.5 L'idonéisme (rapport d'homogénéité entre le problème et la solution)

Dans une des études de cas les plus citées (Sloan-Howitt and Kelling, 1990), les auteurs décrivent une intervention qui est devenue célèbre. On sait que les wagons du métro de New York ont été de 1980 à 1990 si complètement couverts de graffitis qu'on n'en distinguait plus la couleur. Voici comment on réussit à redresser de façon définitive la situation.

Les autorités du métro de New York commencèrent par nettoyer un premier wagon, en appliquant le principe suivant : une fois qu'il avait été lavé, on ne le remettait plus en service s'il avait été à nouveau vandalisé, sans qu'on n'ait encore une fois effacé toutes les déprédations. Si, donc, des vandales recommençaient à couvrir un wagon de graffitis, ceux-ci étaient effacés dès le service terminé, après quoi le wagon ainsi relavé était remis en service. Le public finit par comprendre la détermination des autorités du métro à effectuer le nettoyage des wagons, même si pour parvenir à ce résultat elles devaient commencer par laver inlassablement un premier wagon. Progressivement, les usagers du métro cessèrent leurs déprédations contre les wagons qui avaient été repeints. D'un premier car, on passa à un second et ainsi de suite jusqu'à ce que tout le réseau ait été nettoyé.

On comprendra à quel point la mesure employée était idoine : les cars étant bariolés, on se mit à donc les laver ou à les repeindre. Lorsqu'il y a une telle homogénéité entre le remède et la maladie, la mesure du succès est habituellement facile à effectuer. Dans le cas présent, l'homogénéité entre la maladie et son remède impliquait que tout le processus se passât au sein du *visible* : des cars couverts de graffitis furent lavés et le résultat de cette réfection était visuellement perceptible. Il ne fallait pas être grand clerc pour constater le succès du programme quand le nombre des cars propres excéda celui de ceux qui étaient



encore couverts d'inscriptions. Ce constat est demeuré toutefois trop rudimentaire car il n'explique pas au juste pourquoi la mesure a fonctionné.

On pourrait multiplier ces exemples où le caractère idoine de la mesure rend l'évaluation de son efficacité relativement évidente. Par exemple, le ciblage d'une population non désirée qui s'affiche ouvertement dans un quartier (prostitué(e)s, petits revendeurs de drogue, pseudo-mendiants qui intimident les passants) par une tactique de police intensive produit habituellement la migration perceptible de ces populations. Contrairement à ce qu'on peut anticiper, cette migration ne se résume pas toujours à un simple déplacement vers une zone voisine. Dans le cas de pseudo-mendiants qui se livrent à l'extorsion, il peut se produire une attrition de cette pratique comme telle.

### **3.6 Le cadrage précis**

Cette exigence vaut autant pour les mesures d'intervention que pour l'évaluation de celles-ci. La leçon la plus constante de la mesure du succès de l'intervention policière tient dans l'observation suivante : plus une mesure cible de façon précise un problème et plus elle est susceptible de réussir. C'est ainsi, par exemple, qu'une célèbre expérience conduite à Kansas City a montré que la patrouille automobile qui tournait au hasard sans qu'on lui ait assigné de cible précise était une activité dont la rentabilité était minimale. De façon parallèle, plus une recherche évaluative opère un cadrage précis – par exemple, la résolution d'un problème déterminé – et plus il est facile de l'effectuer. Les grandes recherches américaines qui ont visé à évaluer tout un style d'intervention en lui-même – la « police communautaire » ou la PRP – se sont révélés très coûteuses et leurs résultats sont relativement ambigus.

### **3.7 La multiplication artificielle des sites par la recension des recherches**

Nous avons précédemment vu que l'exigence de validité la plus minimale pour une recherche évaluative était que celle-ci s'effectuât sur plusieurs sites (quand elle ne pouvait même pas s'appuyer sur un concept de recherche qui distinguait entre l'avant et l'après de l'application d'une mesure). Il est dans une certaine mesure possible de remédier à l'unicité d'une expérience sur un seul site par la recension de la littérature de recherche. On peut en effet faire la découverte qu'il se trouve dans les rapports de recherche plusieurs autres sites comparables au sien propre et où des mesures semblables ont été appliquées. Dans ce dernier cas, on multiplie en quelque sorte artificiellement le nombre des sites en leur adjoignant ceux qui sont décrits dans la littérature de recherche. Ce moyen constitue une stratégie ultime de sauvetage, mais ce palliatif peut néanmoins produire des résultats fructueux lorsqu'on découvre dans les recherches antérieures des expériences semblables à celle que l'on veut effectuer et qu'il y a convergence dans leurs résultats. Il va sans dire que cette convergence devra être maintenue lors de l'évaluation de l'expérience à laquelle on se livre soi-même.

### **3.8 La définition du problème**

Nous avons à quelques reprises fait état de l'importance qu'il y avait à définir un problème de sécurité de façon non seulement précise mais innovatrice. C'est bien pourquoi la définition des problèmes fait partie intégrante du processus de leur résolution.

L'exemple canonique à cet égard est celui du métro de New York, que nous avons déjà décrit plus haut. On s'accorde pour dire que le geste décisif a été de définir ce problème d'abord comme un problème d'intendance plutôt que comme un problème de police. La police a parfois été utilisée comme système d'appui pour empêcher que des wagons remis à neuf ne soient à nouveau couverts de graffitis. Cependant, cette pratique n'a pas constitué le cœur de la stratégie. Pour l'essentiel, le succès obtenu dans le métro de New York a été atteint en manifestant de façon vigoureuse et répétée la volonté des autorités compétentes de régler elles-mêmes un problème de délabrement de l'équipement par une stratégie d'intendance qui abordait ce problème de front. Les wagons ayant déjà été vandalisés, il

était trop tard pour remédier à la situation en faisant intervenir la police. Tout au plus aurait-on empêché qu'un graffiti en recouvre un autre, au prix d'une intervention répressive.

Ces illustrations pourraient être multipliées. Par exemple, ce qui était perçu par des commerçants comme un problème d'incivilité d'écoliers a pu être résolu en déplaçant le point où s'arrêtaient les autobus scolaires. Encore une fois, la solution n'était pas de type pénal, bien qu'elle ait été proposée par un policier.

### **3.9 Les évaluations qui reposent sur la distinction de l'avant et de l'après**

Dans sa formulation classique, une stratégie de recherche qui repose sur la distinction entre l'avant et l'après d'une intervention, se réalise en effectuant ce qu'on appelle un pré-test et un post-test. Le pré-test est une mesure quasi-expérimentale des caractères d'une situation problématique avant une intervention de police, alors que le post-test est la mesure de cette situation après l'intervention de la police. Il arrive toutefois souvent que l'on ne puisse suivre de façon rigoureuse cette procédure que constitue l'administration d'un pré-test et celle d'un post-test. Par exemple, lorsque l'intervention de police a débuté un peu avant que le chercheur ne vienne effectuer ses mesures et qu'il lui est impossible d'effectuer un pré-test, au sens rigoureux du terme.

La façon la plus courante de remédier, ne serait-ce qu'en partie, à ce problème est de constituer une série chronologique différenciée dont les pièces constituantes sont des périodes de temps peu étendues. Par exemple, on s'en remettra à la suite des *semaines* plutôt qu'à la suite des *mois* pour suivre l'évolution d'une courbe de délinquance en fonction de celle de l'intensification d'un programme d'action. En réalité, *la constitution de séries chronologiques détaillées, où l'on peut introduire un assez grand nombre de variables, constitue la façon la plus usitée d'effectuer une recherche évaluative, quand on ne dispose pas des conditions optimales ou même convenables pour le faire.*

Finalement, il est possible d'effectuer en cours de recherche ce que nous appellerons un *zoom* temporel au sein d'une série chronologique. Si, par exemple, on soupçonne qu'une évolution intéressante est en train de se passer sur le terrain, on poursuivra la série chronologique en utilisant la journée plutôt que la semaine comme unité temporelle. On pourra même si besoin est distinguer diverses périodes dans une journée.

Voilà donc un ensemble de procédés que l'on peut utiliser pour réaliser des évaluations qui seront peut-être « pauvres » au regard de leurs ressources financières, mais qui atteindront dans une mesure raisonnable les objectifs qu'on peut assigner à la recherche évaluative, quand on renonce au méthodologisme et qu'on cesse de la considérer comme l'une des sciences pures. L'intervention de police, comme l'intervention de sécurité, est affaire de pragmatisme et non d'épistémologie. Ce qui, au demeurant, n'empêche pas que la recherche évaluative produise d'instructives surprises, même quand elle ne se conçoit pas comme une pseudo activité de laboratoire.

Par exemple, on s'est demandé lors de l'évaluation du *Community Policing* à Chicago, pourquoi les programmes avaient du succès dans les quartiers Afro-américains alors qu'ils en avaient très peu dans les quartiers habités par ceux qu'on désigne aux États-Unis comme des « Hispaniques ». On découvrit alors, à la surprise générale, que les « Hispaniques » ne parlaient pas l'anglais, beaucoup d'entre eux ne parlant que l'espagnol, même s'ils habitaient Chicago depuis de nombreuses années (on peut constater le même phénomène à Miami, dont certains quartiers sont exclusivement hispanophones). Comme la police communautaire est en grande partie une police de communication, elle ne peut réussir là où le canal de communication est bloqué.

## **4. Conclusions**

Tout notre exposé a jusqu'ici été consacré à la réalisation d'une évaluation de l'intervention policière dans sa forme. Nous n'avons mentionné qu'au passage les résultats effectifs des évaluations. Nous aimerions présenter en guise de conclusion les principaux résultats des recherches évaluatives.

- **Trois types de résultat** : c'est à tort qu'on pense que les interventions se répartissent en deux catégories, soit celles qui « fonctionnent » et celles qui ne fonctionnent pas. En fait, on trouve une troisième catégorie d'interventions, soit celles qui « dysfonctionnent ». En clair, la recherche évaluative distingue entre trois types d'interventions : (1) celles qui produisent des effets bénéfiques, (2) celles dont les effets sont imperceptibles et (3) celles qui aggravent la situation qu'elles prétendaient résoudre.

- **Les interventions bénéfiques** : nous avons déjà fait allusion à ce qui les différencie, outre leur succès. Ces interventions sont relativement circonscrites et elles sont précisément ciblées. Cette affirmation se vérifie jusque dans le détail. Lawrence Sherman a montré qu'un programme de judiciarisation de la violence conjugale avait des résultats bénéfiques avec des maris qui disposaient d'un emploi stable qu'ils avaient à cœur de protéger mais qu'ils se révélaient dysfonctionnels avec des maris chômeurs qui avaient peu à perdre d'une arrestation par la police et qui, par surcroît, se vengeaient contre leur femme de leur démêlées avec la justice (Sherman, 1992). La signification ultime de ce type de résultat est qu'une stratégie de police se devait avant tout d'être *différenciée* et d'être alignée sur la nature du problème à résoudre. Par exemple, les évaluations ont montré que la poursuite indifférenciée d'une politique de « tolérance zéro » ne produisait pas les résultats escomptés et elle a été dénoncée par celui à qui on en attribue la paternité et qui la considère comme un détournement du *community policing* (Kelling et Cole, 1997; Simonetti Rosen, 1999; Brodeur, 2000 : 189). De façon générale, les interventions produites dans le cadre du *community policing*, qui favorise un style d'intervention peu différenciée, sont évaluées de façon moins favorable que celles qui se produisent dans le cadre de la PRP, plus attentive à la spécificité des problèmes.

- **Les interventions nocives** : les remarques qui précèdent laissent entendre que les interventions indifférenciées et mal ciblées courent le risque du dysfonctionnement. Les recherches sont d'ailleurs parvenues à des résultats plus précis à cet égard : lorsque le seul trait différentiel d'une intervention est son aspect coercitif, elle produit en général une aggravation de la situation. De façon plus particulière, les raids, rafles et autres opérations « coup de poing » manquent leur but et provoquent une dégradation de la situation, surtout quand elles visent sans autre spécification une population jeune (Sherman, 1996 : chap. 8)

- **L'importance de l'accueil** : on sait que les taux d'élucidation des actes de petite et de moyenne délinquance contre les biens sont très bas (Robert, 1999). Ce manque notoire de succès – facilement explicable quand on connaît les limites de l'enquête policière et celles des ressources de la police – se traduit fréquemment dans l'adoption d'une attitude sans empathie par les fonctionnaires de police qui accueillent les plaintes du public. La recherche évaluative a montré que peu de comportements policiers soulevaient autant de frustration parmi les citoyens qu'une indifférence manifeste aux prédatations dont ils sont les victimes. Il faut d'autre part reconnaître qu'il est difficile pour un fonctionnaire de faire de manière continûment convaincante « comme si » la démarche d'un plaignant serait suivie d'un dénouement heureux. En cela comme en plusieurs autres matières, l'activité de police consiste à policer les apparences (Brodeur, sous presse).

Cette dernière observation nous amène à hasarder une prédiction. Il nous semble que l'action sécuritaire sera dans le futur de plus en plus caractérisée par une bifurcation : les organismes publics de police auront tendance à se spécialiser dans la répression de la délinquance de sang et dans les manifestations collectives du crime (criminalité organisée, réseaux terroristes et, à la limite, maintien de l'ordre public); pour ce qui est de la prévention et de la répression de la délinquance contre les biens, ce sera l'apanage du secteur privé, avec l'inégalité sociale qui en résultera. Cette prédiction n'est au vrai pas bien audacieuse, car elle décrit ce qui est maintenant en train de se produire.

## OUVRAGES CITÉS

- Bratton, William (1998), *Turnaround : How America's Top Cop Reversed the Crime Epidemic*, New York : Random House.
- Brodeur, Jean-Paul (sous presse), *L'assemblage policier*, Montréal : Presses de l'Université de Montréal.
- Brodeur, Jean-Paul (2000), « Les dilemmes de la proximité », *Les Cahiers de la sécurité intérieure*, N° 39, 1<sup>er</sup> trimestre 2000.
- Clarke, Ron V. (1992), *Situational Crime Prevention*, New York : Harrow and Heston.
- Goldstein, Herman (1990), *Problem-oriented policing*, New York : McGraw-Hill.
- Johnston, Les (1992), *The rebirth of private policing*, London: Routledge & Kegan Paul
- Kelling, George L. and Catherine M. Cole (1997), *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*, New York : Touchstone.
- McElroy, Jerome E., Coleen A. Cosgrove and Suzan Sadd (1993), *Community Policing: The CPOP in New York*, Newbury Park : Sage.
- Newman, Oscar (1972), *Defensible Space*, Toronto: Collier-MacMillan Canada.
- Ocqueteau, Frédéric, Jacques Frenais et Pierre Varly (2000), *Ordonner le désordre: Une contribution au débat sur les indicateurs du crime*, Paris : La documentation Française.
- Pawson, Ray and Nick Tilley (1994), « What Works in Evaluation Research ? », *British Journal of Criminology*, vol. 34, n° 3, p. 291-306.
- Robert, Philippe (1999), *Le citoyen, le crime et l'état*, Paris : Librairie DROZ.
- Shearing, Clifford D. and Philip Stenning (1985), « From the Panopticon to Disneyworld : the Development of Discipline », dans *Perspectives in Criminal Law: Essays in Honour of John LL.J. Edwards*, A.N. Doob et E.L. Greenspan (dir.), Aurora, Ontario : Canadian Law Books Inc., pp. 335-348. Traduit dans Actes, été 1987, n° 60, pp. 27-33.
- Sherman, Lawrence *et al.* (1996), *Preventing Crime: What Works, What Doesn't, What Promising*, A Report to the United States Congress, Prepared for the National Institute of Justice, College Park: University of Maryland, Department of Criminology and Criminal Justice.
- Sherman, Lawrence (1992), *Policing domestic violence: experiments and dilemmas*, Toronto : Maxwell MacMillan Canada.
- Silverman, Eli B. (1999), *NYPD Battles Crime: Innovative Strategies in Policing*, Boston: Northeastern University Press.
- Simonetti Rosen, Marie (1999), « A LEN interview with Prof. George Kelling, co-author of "Fixing Broken Windows" », *Law Enforcement News*, John Jay College of Criminal Justice/ CUNY, vol. XXV, n° 511-512, p. 8-11 et p. 14.

Skogan, Westley G. and Susan M. Hartnett (1997), *Community Policing, Chicago Style*, New York : Oxford University Press.

Skogan, Westley G. (1990), *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods*, Toronto: Collier Macmillan Canada.

Skogan, Westley G. (1994), « The Impact of Community Policing on Neighborhood Residents: A Cross-Site Analysis », in Rosenbaum, Dennis P., *The Challenge of Community Policing: Testing the Promises*, London: Sage Publications, p. 167-181.

Sloan-Howitt, Maryalice and George L. Kelling (1992), « Subway graffiti in New York City: "Getting up" vs. "Meanin it and cleanin it" », in Clarke, Ronald V., *Situational Crime Prevention*, New York : Harrow and Heston, p. 239-248.

Tufte, Edward R. (1983), *The Visual Display of Quantitative Information*, Cheshire, Connecticut : Graphic Press