

Droit, neurosciences et responsabilité : critique d'une thèse réformiste

Dominic Cliche, *Université Laval*

1. Introduction

Peu de domaines d'investigation philosophique ont échappé au tournant naturaliste qu'a pris la philosophie contemporaine au moins depuis les travaux de Quine sur la naturalisation de l'épistémologie, et la situation n'est pas différente pour les disciplines normatives comme l'éthique et le droit. Suite à cela, certains ont pu se pencher sur ce que signifie un naturalisme juridique ou sur la naturalisation du droit lui-même¹. Dans le présent texte, cependant, mon intérêt se portera sur les conséquences pour le droit d'une forme avancée de naturalisation du comportement humain par le moyen des sciences cognitives et neurosciences.

Si l'on peut taxer d'enthousiasme un peu démesuré ceux qui, face aux imposants succès des sciences de la nature dans le domaine de la production de connaissances, ont pu affirmer que tout savoir devait pouvoir être réduit et formulé dans le langage de ces sciences dures (principalement la physique), force est cependant d'avouer aujourd'hui que ces sciences fournissent au moins un aspect essentiel de notre «meilleure compréhension du monde». En ce sens, tout savoir, tout concept qui se trouverait à ne pas être au moins *compatible* avec les connaissances issues des sciences de la nature devrait être considéré comme fortement improbable, voire être carrément rejeté. Ainsi, toute conception de l'humain qui serait en contradiction avec le fait que ce dernier est un être naturel, une partie de la nature, se trouve exclue de notre meilleure compréhension du monde.

Dans la mesure où le droit vise à réguler et orienter le comportement humain, on serait en droit de s'attendre à ce que les catégories qu'il utilise soient des catégories viables d'un point de vue naturaliste et que, dans la pratique, il traite les humains d'une manière qui ne soit pas incompatible avec notre meilleure compréhension

du monde et de nous-mêmes. Selon certains auteurs, cette attente s'avère nécessairement déçue lorsqu'on tient compte des avancées neuroscientifiques récentes. Notamment, Joshua Greene et Jonathan Cohen, dans leur article « For the law, neuroscience changes nothing and everything »², argumentent que l'explication causale déterministe de l'action humaine qu'offrent les neurosciences nous oblige à abandonner la catégorie de responsabilité criminelle telle qu'elle a été conçue jusqu'à présent et, alors, à délester la justification des sanctions pénales de toute visée de rétribution. Une prise en compte sérieuse des neurosciences exigerait donc, selon eux, de réformer la philosophie sous-jacente au droit criminel³. C'est à l'évaluation de leur argument en faveur de cette thèse que je m'attarderai dans le présent texte. Je commencerai par déterminer quelle forme logique doit prendre une critique radicale du concept de responsabilité criminelle et, ainsi, ce que serait un bon argument permettant de soutenir une thèse réformiste (§2). Au regard de cette analyse, il s'avèrera fort improbable qu'un tel argument puisse être fourni, à moins de modifier l'angle d'approche. C'est ce que font Greene et Cohen dans leur article, et c'est alors vers l'évaluation de leur argument que je me tournerai (§3). Bien que leur tentative semble prometteuse au plan formel, je montrerai qu'elle échoue dans ses prétentions. Je conclurai donc qu'en l'absence d'un argument plus probant en faveur de la thèse réformiste, la responsabilité criminelle n'est pas menacée par les neurosciences (§4).

2. La falsification des catégories juridiques : mode d'emploi

La question est la suivante : la compréhension du comportement humain que peuvent offrir les neurosciences menace-t-elle la responsabilité, au sens où elle est définie par le droit ? Pour pouvoir y répondre de manière efficace, il importe avant tout de se demander ce qui peut compter comme un bon argument permettant de falsifier cette catégorie juridique. La question n'est pas triviale. Le droit n'est pas une théorie de l'esprit, mais un discours normatif à visée pratique, et en ce sens il n'est pas d'emblée évident de savoir comment, si c'est le cas, une théorie de l'esprit ou une conception neuroscientifique du comportement peut venir en falsifier des

composantes. Il n'est pas clair, par exemple, que les propriétés auxquelles font référence les catégories juridiques se laissent saisir par le langage neuroscientifique. De même, ce « problème de traduction » n'implique pas nécessairement pour autant que ces catégories soient incompatibles avec une compréhension naturaliste du monde⁴.

Ainsi, avant de me tourner vers la forme de l'argument, je m'attarderai quelque peu à l'idée que le droit est un discours normatif. J'entends « normatif » ici en deux sens. Le discours juridique est normatif (1) en ce qu'il fournit aux agents des raisons d'agir, c'est-à-dire qu'il a pour visée de réguler et d'orienter l'action des agents, et (2) en ce qu'il mobilise des catégories définies elles-mêmes sur la base de considérations normatives. L'affirmation (1) n'est pas celle qui m'occupera principalement. Elle n'est pratiquement pas controversée, du moins depuis Hart⁵, et reprend directement l'idée que le droit s'adresse et s'applique à des « personnes », c'est-à-dire à des êtres capables d'exercer leur raison pratique (*practical reasoners*)⁶.

L'affirmation (2) est plus pertinente à mon propos, mais elle peut s'avérer plus problématique et, surtout, plus ambiguë. Elle repose sur deux postulats. Premièrement, le droit ne décrit pas des processus naturels, mais formule plutôt des catégories artificielles ainsi que les critères permettant de les appliquer, c'est-à-dire des critères permettant de déterminer si un certain comportement ou un certain individu peut être rangé sous une de ces catégories. Deuxièmement, et c'est ce qui constitue l'aspect proprement normatif des concepts ou catégories juridiques, les critères de leur application sont justifiés au regard de considérations normatives, c'est-à-dire des principes appartenant à un ensemble de règles dont la visée constitutive est de réguler les comportements, soit d'*approuver* les comportements *acceptables* et de *punir* les comportements *inacceptables*. Par exemple, aucun acte n'est « criminel » en soi ; pour appliquer cette catégorie à des faits, il faut au préalable que des critères aient été formulés pour discriminer un acte juridiquement acceptable et un acte juridiquement inacceptable. Pour ce faire, il est impossible de sortir complètement du vocabulaire normatif.

Il y a donc à l'intérieur du droit une certaine indépendance du normatif par rapport à la description empirique, dans la mesure où, par exemple, une fois que les faits empiriques sont décrits de manière neutre, il demeure à se demander s'ils constituent une infraction criminelle, c'est-à-dire un comportement inacceptable du point de vue du droit criminel. Les concepts juridiques, cependant, sont *aussi* nécessairement sensibles aux faits, au sens où ils comportent une composante empirique qui peut être vérifiée ou falsifiée. Un crime n'est jamais *seulement* une suite d'événements décrits de manière neutre, mais appliquer la catégorie «crime» demande toujours une analyse attentive et scrupuleuse des faits auxquels s'appliquent les critères.

Cette présentation assez sommaire de l'aspect normatif du discours juridique permet d'avancer deux propositions :

(2.1) La définition des catégories juridiques normatives n'est pas *a priori* contrainte par les faits. La définition est toujours plutôt le résultat d'une prise de position sociale, au regard de principes ou d'engagements normatifs préalables, par rapport à ce qui est acceptable ou non. Ainsi, seul un argument contenant au moins une prémisse normative semble pouvoir mettre en danger une de ces catégories. En ce sens, une catégorie juridique ne peut pas être démontrée fausse simplement sur la base de données empiriques, mais elle peut être démontrée *injustifiée* d'un point de vue normatif.

(2.2) L'application des catégories, elle, est contrainte par les faits. S'il est possible de falsifier une catégorie juridique, cela devra être fait indirectement, en montrant que la catégorie est *inapplicable*. Il s'agit de voir que si la définition d'une catégorie juridique est avant tout normative, cela n'empêche pas que l'extension de cette catégorie puisse s'avérer *vide*. On peut bien définir une «personne» de telle ou telle manière, si on découvre que, dans les faits, personne ne satisfait aux caractéristiques sélectionnées dans la définition, alors on a une catégorie bien claire, mais sans effet. Par exemple, on pourrait bien accorder des droits spécifiques aux extra-terrestres, mais, jusqu'à preuve du contraire, les catégories auxquelles feraient référence ces droits ne pourraient jamais être invoquées dans la décision d'un tribunal. Ainsi, si les caractéristiques pertinentes

sont sélectionnées normativement, il faut tout de même que ces caractéristiques existent. Le droit n'a d'autorité que sur la pertinence, et non sur l'existence.

Un argument permettant de falsifier une catégorie juridique se présente alors, au point de vue formel, comme la conjonction de deux prémisses, l'une qui énonce la définition de la catégorie, selon (2.1), et l'autre qui formule un jugement d'existence concernant certaines ou toutes les propriétés empiriques identifiées dans la définition, selon (2.2) :

Tout ce qui est P a les propriétés empiriques $\langle x,y,z \rangle$;
Il n'existe rien qui ait les propriétés empiriques $\langle x,y,z \rangle$;
Donc, il n'existe rien qui soit P, i.e. l'extension de P est vide.

Dans le cas où P est une catégorie juridique, la conclusion est assurément indésirable, au sens où elle fait en sorte que la catégorie est inapplicable. Or c'est une attente légitime, voire nécessaire, qu'une catégorie juridique puisse être appliquée, sans quoi elle ne peut pas remplir sa visée pratique et orienter l'action des agents ou les décisions des officiers. C'est l'ajout d'une prémisse normative supplémentaire, soit « l'extension de P ne *doit* pas être vide »⁷, qui fait que l'argument peut proprement compter comme une réduction à l'absurde. Dans la mesure où la vérité de la prémisse mineure (le jugement d'existence) repose sur un procédé de vérification indépendant, on ne peut pas tenter d'éviter la contradiction simplement en la rejetant, bien que cela puisse inciter à retourner dans le laboratoire. Il semble alors que le problème repose sur le fait que la prémisse majeure (la définition) est inadéquate, que la catégorie juridique est problématique et que ce soit elle qui doive être révisée ou abandonnée.

Revenons au cas particulier de la catégorie de responsabilité criminelle. Pour formuler un argument qui permettrait de la falsifier, on doit premièrement définir ce qui fait qu'un individu puisse être tenu responsable d'une infraction quelconque, de manière à déterminer si cette définition implique l'existence de propriétés ou de caractéristiques falsifiables empiriquement. De manière générale, les critères déterminant la responsabilité découlent de la proposition

(1) présentée ci-dessus, comme quoi le droit est normatif en ce qu'il fournit aux agents des raisons d'agir, et donc qu'il s'applique à des «personnes», c'est-à-dire à des êtres capables d'exercer leur raison pratique (*practical reasoners*). Ainsi, un individu est tenu responsable d'une action qu'il a posée seulement s'il possédait, au moment de l'action, les capacités normales et nécessaires pour obéir à la loi, à laquelle il a eu une «possibilité équitable» (*fair opportunity*) d'obéir⁸. De manière plus spécifique, cet individu capable, volontaire et conscient doit avoir posé le geste qui lui est reproché (preuve de l'acte matériel de culpabilité, ou *actus reus*). Il doit aussi l'avoir fait avec l'intention qui lui est reprochée (preuve de l'intention criminelle, ou *mens rea*)⁹. Pour être tenu responsable, un individu doit donc jouir de certaines capacités, dont celles de former des intentions et d'assurer un degré suffisant de contrôle sur son action, c'est-à-dire de soumettre ses états motivationnels à des raisons à un niveau jugé suffisant.

Cependant, certaines restrictions s'appliquent. Dans le cadre du droit criminel canadien, il est reconnu qu'un individu n'est pas criminellement responsable si un trouble mental le rend, au moment de commettre l'acte, «incapable de juger de la nature et de la qualité de l'acte [...], ou de savoir que l'acte [...] était mauvais¹⁰». Pour être tenu responsable, il doit savoir ce qu'il fait, avoir des croyances minimalement rationnelles ou ne pas être déconnecté par rapport à la réalité. Un exemple figurant dans la jurisprudence récente est celui de Pascal Morin, reconnu non criminellement responsable du meurtre de sa mère et de ses deux nièces, suite à un épisode de schizophrénie où il était convaincu qu'elles étaient possédées du démon et que lui, qui se prenait pour Jésus, devait les tuer pour ainsi sauver la planète¹¹. Il aurait eu beau mettre en œuvre une rationalité moyens-fin impeccable, cela n'est pas suffisant pour qu'il soit tenu criminellement responsable de son geste. Il en est de même pour un individu qui aurait agi sous l'effet d'une contrainte qui, dans la situation particulière où il se trouvait, s'avérait être une meilleure raison d'agir que la loi qu'il a transgressée¹². Par exemple, s'il a volé sous l'effet d'une menace de mort, on considérerait qu'il était plus raisonnable de commettre l'infraction que de mourir.

Ainsi, une catégorie de responsabilité applicable inclurait les éléments suivants : la personne tenue responsable a la capacité de former des intentions et la capacité de soumettre ses états motivationnels à des raisons, elle fait exercice de ces capacités au moment de l'action, elle n'est pas déconnectée par rapport à la réalité et agit en l'absence de contraintes excessives¹³. Une réfutation empirique devra montrer que personne, ou qu'un nombre trop important de personnes, ne peut être tenu suffisamment rationnel ou libre pour que lui soit appliquée la catégorie de responsabilité.

Une *bonne* réfutation empirique devra cependant tenir compte d'une contrainte supplémentaire. En effet, les éléments retenus peuvent mener à deux interprétations, l'une d'entre elles étant trop large pour pouvoir servir, dans le cas présent, comme prémisse dans un argument fondé. Les deux interprétations se distinguent quant au rôle que joue l'explication causale dans la disculpation. Dans la première, un motif valable de disculpation peut consister en le fait que le comportement de l'agent soit déterminé par des causes, externes à sa vie mentale, qu'il ne pouvait contrôler. Dans l'autre, seules certaines causes externes peuvent jouer ce rôle, et uniquement dans la mesure où elles sont significativement liées à une incapacité spécifique. L'adoption de la première interprétation est ce que Stephen J. Morse a nommé « l'erreur psycholégale fondamentale » (EPF ; *fundamental psycholegal error*). Par exemple, une tumeur au cerveau peut être une cause véritable rendant mon comportement fort étrange, mais elle ne m'excuse pas de tout écart de conduite, à moins qu'on puisse montrer qu'elle m'empêche d'exercer ma raison pratique à un niveau suffisant. Par exemple, qu'elle me rende plus agressif qu'à l'habitude n'implique pas qu'elle ruine ma capacité à exercer un contrôle minimal sur mon agressivité.

The discovery that the brain, including a brain abnormality, played some causal role in the production of what is undeniably human action does not lead to any legal conclusions about responsibility. The proper internal question is whether the neuroscience evidence helps to establish the presence or absence of action, mental states or a genuine affirmative defense, such as lack of rational capacity¹⁴.

L'exclusion de l'EPF ne rend pas toute falsification impossible, bien sûr, mais elle introduit une contrainte conceptuelle supplémentaire sur la définition de la responsabilité. Cela fait en sorte qu'un *bon* argument devra montrer *directement* qu'un nombre important d'êtres humains n'ont pas les capacités minimales nécessaires pour être considérés comme des êtres pouvant exercer leur raison pratique et ce, peu importe la vérité du déterminisme causal qu'impliquent les sciences de la nature.

Une telle démonstration est improbable. Le meilleur candidat, il me semble, est une forme de matérialisme éliminativiste (ME)¹⁵. Une telle position requiert l'abandon du vocabulaire de la psychologie naïve (*folk psychology*), c'est-à-dire principalement des termes dénotant des attitudes propositionnelles (désirs, croyances, intentions, etc.) nécessaires à la formulation de la conception minimale de la raison pratique qui est utilisée par le droit. ME invite à réformer en profondeur la compréhension qu'ont d'eux-mêmes les agents, et ainsi semble priver nécessairement le droit du cadre général de catégorisation du comportement qui en est constitutif. Cependant, même une proposition aussi radicale n'a peut-être pas de conséquences aussi catastrophiques pour le droit qu'il n'y paraît. Notamment, Paul Churchland affirme :

Eliminative materialism thus does not imply the end of our normative concerns. It implies only that they will have to be reconstituted at a more revealing level of understanding, the level that a matured neuroscience will provide¹⁶.

ME doit maintenir un certain sens des termes «rationalité» et «normativité» à travers la traduction dans le langage des neurosciences, ne serait-ce que pour assurer sa cohérence en tant que théorie qui vise à être rationnellement justifiée. Ainsi, à moins de rejeter en entier le droit en tant que système normatif, ME n'affecte pas la catégorie de responsabilité en elle-même ; il ne fait que rendre abusivement compliquée la description des capacités requises¹⁷. Dans la mesure où l'on peut continuer à discriminer entre les processus selon ce que l'on valorise ou encourage – la définition

des catégories juridiques – et dans la mesure où cette discrimination peut au moins en principe avoir un impact pratique sur les processus qui ont effectivement cours – ce qui a trait à la normativité ou au caractère pratique des catégories, incluant la capacité pour les cerveaux de tenir compte des catégories pour l'action –, on a les éléments essentiels pour constituer le cadre général et formel du droit en tant que système normatif de régulation des comportements.

3. Une tentative prometteuse de falsification : Greene et Cohen sur le rôle des intuitions morales

Pour falsifier la catégorie juridique de responsabilité, il faut d'abord définir les critères normatifs de la responsabilité. Une telle définition peut être justifiée ou injustifiée au regard de considérations normatives, mais elle n'est pas directement falsifiable. Ensuite, il faut déterminer si la catégorie est applicable en déterminant quelles propriétés ou caractéristiques empiriques sont impliquées par la définition et si elles existent réellement. Pour obtenir une conclusion négative, compte tenu de l'état actuel du droit, il faudrait montrer qu'un nombre important de personnes ne peuvent pas être considérées comme des êtres exerçant de manière suffisante leur raison pratique. Or, à moins de nier que le droit s'applique à des personnes à qui il fournit des raisons d'agir, une caractéristique essentielle au droit lui-même, il est impossible de montrer que la catégorie juridique de responsabilité est vide. L'éventualité d'une telle falsification s'avère donc fort improbable.

En ce sens, pour le droit, les neurosciences ne changent rien. Cependant, Greene et Cohen¹⁸ affirment que ce constat est fragile et que, dans l'avenir, les neurosciences sont en fait appelées à tout chambouler. Leur thèse principale est que l'application des catégories juridiques n'est pas déterminée par leur seule définition explicite, mais aussi par les intuitions morales implicites qui les justifient : «The legitimacy of the law itself depends on its adequately reflecting the moral intuitions and commitments of society. If neuroscience can change those intuitions, then neuroscience can change the law¹⁹». Dans le cas de la responsabilité criminelle, ils affirment que cette catégorie juridique peut être modifiée en fonction des intuitions

morales des individus concernant la responsabilité et le libre arbitre. Or ces intuitions morales imposent des critères plus exigeants que la catégorie formulée par le droit. L'intuition à laquelle ils font appel est en fait une réhabilitation de l'EPF, comme quoi un agent ne peut légitimement être considéré responsable d'une action que s'il est la cause ultime (CU ; *uncaused causer*) de cette action. Il serait considéré immoral de punir quelqu'un qui ne satisfait pas à ce critère. Ainsi, une fois que les neurosciences auront rendu manifeste que l'action humaine est soumise au déterminisme causal, les individus devraient devenir réfractaires à tenir quiconque responsable de ses gestes, et la catégorie de responsabilité criminelle deviendra alors inapplicable.

C'est cet argument que je souhaite évaluer dans la présente section. Je commencerai par justifier le changement qu'ils opèrent dans l'angle d'approche et montrer ce que cela implique quant à l'argument proposé en faveur de la thèse réformiste (§3.1), pour ensuite analyser et critiquer leurs deux arguments principaux : une expérience de pensée (*Mr Puppet*) sensée isoler l'intuition en question (§3.2) et l'explication de cette intuition d'un point de vue évolutionniste (§3.3). Dans les deux cas, je montrerai que leur démonstration ne permet pas de soutenir leur position réformiste.

3.1. Un nouvel argument tenant compte du rôle des intuitions morales

Le but de Greene et Cohen est de montrer que l'EPF est en fait fortement ancrée dans les intuitions morales des individus, qu'elle n'est finalement pas une erreur, et qu'elle fait ainsi pression sur la définition de la catégorie juridique de responsabilité. Cela fait en sorte que la falsification de la catégorie de responsabilité serait possible simplement sur la base du fait que les agents sont soumis au déterminisme causal impliqué par les sciences de la nature, et principalement qu'ils sont déterminés par des causes neurologiques extérieures à leur vie mentale consciente.

Ce changement dans l'angle d'approche implique une différence importante par rapport à l'argument que j'ai présenté plus haut. Cette différence repose dans le fait que Greene et Cohen s'intéressent à l'impact de croyances générées par des intuitions,

et non seulement aux catégories positives du droit et aux états de fait en eux-mêmes. Les résultats scientifiques *en tant que* résultats scientifiques ne sont que peu importants dans l'argument : ce sont leurs effets, lorsqu'intégrés par des individus raisonnables parmi leurs croyances et mis en relation avec leurs intuitions, qui sont importants. L'argument devient alors le suivant :

Toute personne a l'intuition morale que pour être légitimement tenu responsable d'une action, un agent doit être la cause ultime de cette action ;

Les neurosciences rendent manifeste qu'aucun agent n'est la cause ultime de quelque action que ce soit²⁰ ;

Donc, toute personne consciente de ce dernier fait croit qu'aucun agent ne peut être légitimement tenu responsable de quelque action que ce soit.

Cette différence semble faire en sorte que l'argument de Greene et Cohen prête le flanc à la critique. En effet, pourquoi s'intéresser à ce que les individus *croient* être le cas, et non pas directement à ce qui *est* le cas ? Les croyances des individus pourraient bien être *fausses*, auquel cas le détour n'en vaut pas la chandelle. Le droit aurait raison de ne pas prendre en compte une intuition qui mène à des croyances fausses, et on peut espérer que les arguments des philosophes compatibilistes²¹ puissent en venir à corriger ces intuitions. Ainsi, on pourrait affirmer que l'argument de Greene et Cohen échoue, qu'il ne règle pas la question, mais plutôt la contourne, en remplaçant, par une acceptation non critique des intuitions naïves des individus, ce qui serait véritablement une bonne réponse à la question et qui ne pourrait se passer d'arguments métaphysiques.

Cependant, il me semble qu'une telle critique n'est pas fondée, dans la mesure où elle néglige ce qui fait justement une force de l'argument de Greene et Cohen, soit le fait de prendre au sérieux le caractère normatif des catégories juridiques. Si l'on accepte que la responsabilité n'est pas une catégorie naturelle, mais une construction humaine visant certains objectifs de régulation des comportements – par exemple, réguler l'attribution du blâme et des louanges de manière juste entre les individus – alors il semble que

pour bien la définir, on doit s'attarder à décrire la pratique humaine qui consiste à tenir un agent pour responsable de ses actions. Les intuitions des agents concernant à ce qui est requis par cette pratique sont vraisemblablement des plus importantes, voire essentielles, à une telle description.

Une autre critique possible s'appuie sur le fait que le droit n'a pas retenu l'intuition morale CU pour la définition de la responsabilité; cette intuition n'est donc pas du droit. Pourquoi le droit devrait-il alors s'en soucier, surtout si elle est métaphysiquement problématique? La réponse de Greene et Cohen est qu'en fait, le droit a déjà intégré implicitement CU, quoique sans que ne transparaisse d'emblée la tension que cela implique. Un agent responsable doit minimalement pouvoir exercer sa raison pratique; or on ne peut pas comprendre ce que signifie le concept d'agent pouvant exercer sa raison pratique sans référer à CU (cet argument sera explicité en §3.3). Ce que cela implique, c'est que personne ne serait prêt à considérer quiconque comme responsable sans qu'il ne satisfasse à CU. Ainsi, lorsqu'on applique la catégorie de responsabilité, c'est toujours à un comportement qui semble satisfaire à CU, peu importe ce qu'en dit explicitement le droit; face à un comportement qui ne semble pas y satisfaire, la catégorie est jugée ne pas s'appliquer, ou du moins ne pas s'appliquer de manière légitime (cet argument sera explicité en §3.2). Les questions qu'on pose pour déterminer si un agent est responsable, s'il est suffisamment rationnel, s'il mérite une punition, disent Greene et Cohen, sont toujours motivées par une conception de la raison pratique plus riche que celle explicitement fournie par le droit.

En résumé, pour les deux auteurs, il y a une différence entre ce que le droit *dit* et ce qu'il *présuppose* pour gagner l'appui des citoyens – pour que ces derniers acceptent sa normativité et pour qu'il leur apparaisse mener à des décisions légitimes. La catégorie de responsabilité est appliquée sur la base de CU et lorsqu'une tension apparaît, dans une situation limite, entre la catégorie juridique et la catégorie plus riche qui motive réellement les agents, la catégorie juridique semble nécessairement erronée et inapplicable sur une base normative. En l'absence d'une justification morale de la notion de

responsabilité, il semble que la catégorie juridique ne puisse plus être considérée comme *socialement légitime* (bien que toujours *légalement valide*). C'est cette «délégitimisation» de la catégorie de responsabilité qui appelle avant tout à une réforme du droit.

3.2. *Mr Puppet*

L'argument proposé par Greene et Cohen est valide ; il reste à déterminer s'il est fondé. Il est fondé si l'intuition morale à laquelle ils font appel, soit qu'un agent ne peut être considéré responsable d'une action que s'il est la cause ultime (*uncaused causer*) de cette action²², est partagée au sein de la population. Pour appuyer leur thèse, Greene et Cohen proposent une expérience de pensée, celle de *Mr Puppet*, visant à isoler l'intuition en question. Cependant, si dans les faits plusieurs intuitions conflictuelles pouvaient être isolées, un problème surviendrait nécessairement pour leur argument. Dans un tel cas, la question normative de savoir quel concept de responsabilité on devrait adopter demeurerait ouverte. C'est, je crois, ce qui fait que l'argument de Greene et Cohen ait, au final, peu de chance de succès.

L'expérience de pensée est la suivante²³. Imaginons qu'un groupe de scientifiques parvenait, par la manipulation génétique et un contrôle accru des variables environnementales, à produire un individu (*Mr Puppet*) de manière à ce que chacun de ses désirs et chacune de ses croyances, et ainsi chacun de ses comportements intentionnels, soient le résultat nécessaire de la planification effectuée par le groupe de scientifiques. Ces scientifiques, dans la mesure où ils réussissent, peuvent prédire chaque fait et geste de *Mr Puppet*. Plus encore, ils peuvent se considérer comme les véritables agents de la vie de *Mr Puppet*. Ils l'ont créé justement pour qu'il soit tout ce qu'il est et accomplisse tout ce qu'il accomplit. Un jour, *Mr Puppet* commet un crime qui, nécessairement, est aussi un comportement planifié par le groupe de scientifiques.

Mr Puppet agit selon ses intentions, formées selon ses désirs et ses croyances, en toute conscience et sans coercition excessive. Il devrait donc, à la lumière des critères fournis par le droit, être tenu responsable de ses actes. Cependant, Greene et Cohen affirment, à raison il me semble, que de le tenir ainsi responsable est contre-

intuitif. Plutôt, on est tenté de le disculper complètement de son crime, au moins au sens où on ne considère pas qu'il *mérite* d'être puni pour ce crime. Pour expliquer le manque de concordance entre la conclusion tirée à partir de la catégorie juridique et celle tirée intuitivement, ils suggèrent que la question qu'on se pose dans les faits n'est pas de savoir s'il a agi intentionnellement, mais s'il a agi librement, si c'était bien *lui* qui a agi. Ce qui pose problème, c'est qu'il apparaît être le jouet de forces qui le dépassent, d'influences causales qu'il ne peut aucunement contrôler, et donc qu'il ne peut pas être considéré comme étant la cause ultime de ses actions. Dans la mesure où les neurosciences – du moins lorsqu'elles seront parvenues à maturité – offrent une telle description déterministe du comportement de tout agent, la conclusion qu'on tire intuitivement dans le cas de *Mr Puppet* devra s'étendre à toutes les actions et tous les agents.

Cependant, il me semble que l'expérience de pensée ne permet pas réellement d'isoler l'intuition qui soutiendrait CU. Le problème principal est que Greene et Cohen sautent trop rapidement de l'intuition d'iniquité déclenchée par l'application de la catégorie de responsabilité criminelle à *Mr Puppet* et de la question sous-jacente concernant la légitimité de l'attribution de l'action, à une explication de cette intuition et de cette question sur la base d'une conception du libre arbitre entendu comme la capacité d'être la cause ultime de ses actions. Il y a pourtant des explications alternatives possibles. L'intuition d'iniquité est distincte de l'intuition sur le libre arbitre qu'ils proposent pour l'expliquer, et ils doivent pouvoir montrer que les autres explications possibles ne sont pas crédibles. C'est-à-dire qu'ils doivent présenter un argument supplémentaire pour trancher en faveur de leur interprétation de l'histoire de *Mr Puppet*. Le reste de cette section sera consacrée à une justification de cette objection.

Tout d'abord, l'histoire de *Mr Puppet* est présentée dans une narration qui dépasse le strict déterminisme causal. Pour rendre compte d'un ensemble déterminé de causes qui mènent à une certaine action, on reconstruit une histoire où cet ensemble déterminé de causes est préalablement et soigneusement, *consciemment*, planifié pour mener à cette action. Du point de vue de *Mr Puppet*, cette

situation a de forts relents de fatalisme, soit que ce qui arrive *doit* arriver, peu importe qu'il le veuille ou non. Cela est cependant une thèse plus forte que le déterminisme et il n'est pas déraisonnable de penser que des intuitions distinctes seraient déclenchées si le scénario présenté en était un où il n'est fait mention que de forces purement impersonnelles, sans aucune explication téléologique. Que l'on considère que *Mr Puppet* ne soit pas suffisamment libre pour être tenu responsable, cela ne repose pas nécessairement sur le fait que la suite des causes soit déterminée (au sens du déterminisme naturel), mais peut-être sur le fait qu'elle soit soigneusement planifiée *par un autre agent*. Greene et Cohen, bien sûr, attestent de cette différence, mais ils la tiennent pour non significative.

Pourtant, cette différence pourrait s'avérer significative et Greene et Cohen devraient se justifier davantage si une interprétation alternative permettait de rendre compte de la différence entre les conclusions juridiques et intuitives sans faire appel à une conception du libre arbitre comme capacité d'être la cause ultime de ses actions²⁴. Une manière d'expliquer l'intuition d'iniquité tout en demeurant neutre quant à CU pourrait être de déterminer les influences causales qui ont joué un rôle dans la formation des intentions et des croyances qui ont mené à l'action, pour ensuite déterminer, selon un seuil que l'on considère raisonnable, lesquelles sont acceptables pour qu'un agent ne cesse pas de considérer que l'action en question est *son* action. On peut, par exemple, formuler le principe contrefactuel suivant : « Une action *a* est attribuable à un agent *A* si, et seulement si, *A* la considérerait comme étant bel et bien *son* action même s'il connaissait les influences causales pertinentes l'ayant mené à accomplir *a* ».

Un tel principe semble pouvoir fournir les résultats escomptés, c'est-à-dire discriminer explicitement ce qu'on est porté à discriminer de manière relativement intuitive, entre un agent responsable de *a* et un agent non responsable de *a*. L'explication fait appel, non pas à la notion de causalité elle-même, mais à certains types de causes qui sont mobilisées et qui posent problème pour la responsabilité ou l'authenticité. On peut accepter bien des influences causales sur lesquelles on n'a aucun contrôle tant qu'on ne juge pas qu'elles ne

devraient pas nous influencer. Par exemple, je n'ai aucun problème à admettre que mon caractère est déterminé par mes gènes et par des expériences que j'ai eues au fil de ma vie; je pourrais cependant voir un problème dans le fait que mes parents aient fait modifier mes gènes pour que je sois, par exemple, un bon garçon obéissant.

Pour revenir finalement à *Mr Puppet*, je proposerais que c'est en raison de l'identification de causes *inacceptables* du point de vue de l'authenticité de l'action qu'on est porté à trouver injuste de le tenir responsable ou de le punir pour son geste. D'ailleurs, le fait que l'histoire ne précise pas si *Mr Puppet* sait qu'il est manipulé ou non me semble significatif. L'intuition d'iniquité me semble plus forte dans l'éventualité où *Mr Puppet* ignore qu'il est manipulé. S'il le sait et l'accepte, au contraire, on est moins enclin à trouver son sort injuste, dans la mesure où il aura *fait siens* les motivations et les comportements qui résultent de sa condition. Encore là, un sentiment d'iniquité persistant pourrait être imputé au fait, par exemple, qu'on juge qu'aucun individu raisonnable ne puisse consentir à une telle chose. Un critère objectif pourrait ainsi être préféré au critère subjectif présenté plus haut, en utilisant la norme de la personne raisonnable: «Une action *a* est attribuable à un agent *A* si, et seulement si, une personne raisonnable, dans la situation de *A*, la considérerait comme étant bel et bien *son* action même si elle connaissait les influences causales pertinentes l'ayant mené à accomplir *a*».

Greene et Cohen pourraient répliquer que rien dans ce que j'avance ne garantit que je ne présuppose pas implicitement une forme de libre arbitre comme celle qu'ils attaquent. En effet, rien ne le garantit. Mais rien ne garantit l'inverse non plus, soit que ce n'est pas le type de cause en question qui déclenche mon intuition. Dans tous les cas, un argument supplémentaire est nécessaire pour trancher la question, d'un côté comme de l'autre.

Si mon interprétation alternative s'avérait, cela impliquerait malgré tout qu'il faudrait compléter la catégorie juridique de responsabilité pour qu'elle en tienne compte, car le système judiciaire actuel ne considère pas ce genre d'origine causale comme une source potentielle de disculpation. Par exemple, Sifferd²⁵ montre que le droit demeure mal outillé pour bien gérer ce qui peut constituer une défense

acceptable face à la possibilité d'implantation de désirs chez un autre individu par le moyen des nanotechnologies, un cas relativement similaire à celui discuté ici. C'est d'ailleurs pourquoi je crois que l'argument de Greene et Cohen semble fonctionner; le droit actuel devrait effectivement traiter de la même manière *Mr Puppet* et son homologue non manipulé. Cela n'est pas à dire qu'aucune révision mineure ne puisse remédier au problème. Cependant, si Greene et Cohen avaient raison, la catégorie de responsabilité ne pourrait pas être complétée sans en compromettre l'application et entrer en conflit avec notre meilleure compréhension du monde. Le problème repose ainsi strictement sur l'introduction de la notion de libre arbitre comme capacité pour l'agent d'être la cause ultime de ses actions. Il est néanmoins possible d'en faire l'économie et de s'en tenir au fait que, collectivement, nous jugeons inacceptable qu'un individu puisse être aussi fondamentalement manipulé par d'autres. Cela ne va pas à l'encontre de notre meilleure conception du monde. Il serait alors possible d'espérer que le niveau raisonnable d'intervention extérieure puisse être déterminé normativement par une cour et ce, sans compromettre l'application de la catégorie de responsabilité. En somme, l'exemple de Greene et Cohen montre que la catégorie de responsabilité qu'on applique intuitivement est plus exigeante que celle explicitement fournie par le droit, mais, considérant ce que j'ai avancé, la catégorie appliquée n'est pas nécessairement *trop* exigeante au regard de notre meilleure compréhension du monde.

3.3 L'explication évolutionniste des intuitions

L'expérience de pensée de *Mr Puppet* visait à isoler l'intuition morale sous-tendant CU. Or j'ai montré qu'une telle intuition ne pouvait pas être isolée avec certitude, et donc que la conclusion de Greene et Cohen ne découle pas nécessairement des prémisses. Greene et Cohen peuvent cependant offrir un autre argument, à partir d'une explication évolutionniste des intuitions. Comme Greene et Cohen considèrent avoir démontré leur thèse, l'argument évolutionniste qu'ils offrent vise à expliquer pourquoi on ne considère comme un agent véritable que celui qui satisfait à CU. Cependant, dans la mesure où il n'est pas clair qu'on ait l'intuition en question,

l'argument peut tout de même servir une autre fonction: celle de montrer qu'on a cette intuition, et non une autre, qu'on a évolué de telle manière qu'on a bel et bien cette intuition. En ce sens, il peut servir à compléter, et non simplement expliquer, la thèse avancée par Greene et Cohen. C'est cette dernière fonction qui m'intéressera ici.

L'explication qu'ils mettent de l'avant consiste à montrer que l'humain a eu un avantage évolutif à développer deux systèmes cognitifs permettant de traiter différemment les objets de son environnement selon deux catégories distinctes²⁶. Un premier système, celui de la «physique naïve» (*folk physics*) traite des objets inanimés et leur attribue intuitivement certaines propriétés. Un second système, celui de la «psychologie naïve» (*folk psychology*) traite des objets animés, en leur attribuant aussi certaines propriétés, soit les «unseen features of minds: beliefs, desires, intentions, etc.²⁷». Plus encore, Greene et Cohen avancent que la caractéristique principale de la psychologie naïve, et celle qui permet de la distinguer de la physique naïve, est que les objets qu'elle traite semblent se mouvoir non pas en raison de causes physiques, mais plutôt en raison de buts qu'ils poursuivent. Ainsi, ils suggèrent qu'une caractéristique essentielle d'un esprit (*mind*) est d'être la cause ultime de ses actions, ce qui validerait CU. De plus, dans la mesure où l'on ne blâme dans les faits que les objets/agents que traite la psychologie naïve, «[t]o see something as a responsible moral agent, one must first see it as having a mind. But, intuitively, a mind is, among other things, an uncaused causer²⁸».

Selon notre meilleure conception, le monde dans lequel on vit est déterminé par les lois de la nature. En ce sens, si la description offerte par Greene et Cohen est juste et que de considérer quelqu'un (ou quelque chose) comme un agent implique toujours intuitivement de le voir habité par un esprit, et conséquemment de le considérer comme la cause ultime de ses actions, alors il n'existe pas d'agents – encore moins d'agents responsables. Cependant, cette description s'avère problématique sur un aspect crucial: CU ne joue pas le rôle central qu'il est supposé y jouer.

En effet, CU pourrait – et devrait – servir à distinguer un objet tel qu'il est traité par la psychologie naïve d'un autre objet tel qu'il

est traité par la physique naïve, ou alors servir de lien logique entre la description en termes de psychologie naïve et l'attribution de responsabilité morale. Or il semble, d'une part, tout à fait plausible, à partir de l'information que donnent Greene et Cohen, que ce qui distingue un agent d'un objet inanimé ne soit pas tant qu'il soit considéré comme la cause ultime de ses actions, mais plutôt que les causes pertinentes dans la compréhension de son action soient des désirs, des croyances et des intentions ; en somme, que son action soit médiatisée par un *certain type de cause* – des événements mentaux qui peuvent en retour être causés par des événements physiques ou neurologiques – plutôt que d'avoir son impulsion première à l'extérieur de la chaîne causale naturelle²⁹. D'autre part, le même genre d'objection permet de montrer que le détour par CU n'est pas nécessaire pour rendre compte de l'attribution de blâme. Greene et Cohen donnent pour exemple qu'on ne blâme pas la pluie qui tombe, mais qu'on blâme le voisin s'il décide d'arroser nos convives lors d'une fête estivale. Cependant, la différence fondamentale entre la pluie et le voisin n'est pas nécessairement que ce dernier est considéré être la cause ultime de ses actions. Par exemple, le voisin peut se voir attribuer des intentions, des buts ; la pluie, non. Du moment qu'un comportement peut être décrit dans le langage téléologique de la raison pratique, il est susceptible d'être attribué à une personne qui en sera tenue responsable. Évidemment, d'autres critères pourront s'ajouter pour rendre compte de l'attribution du blâme, comme par exemple que le voisin n'agit pas sous la contrainte, qu'il ne pose pas ce geste pour sauver nos convives d'un feu qui les menace, ou pour éviter un quelconque plus grand mal, etc. Néanmoins, le fait d'être *a priori* admissible à l'attribution de la responsabilité et du blâme ne requiert pas de référence à CU. Si la pluie se voyait attribuer des états mentaux et poursuivait des buts, dont celui de nous embêter, il serait probablement intuitif de la blâmer pour avoir gâché notre fête.

Ainsi, en l'absence d'un argument supplémentaire, il semble que Greene et Cohen en soient réduits à présupposer que le compatibilisme est faux, ou du moins intuitivement faux, alors que les données ne permettent pas, pour l'instant, de trancher la question. Ne pouvant pas fournir une base plus solide à CU au moyen de leur

explication évolutionniste, force est d'admettre que la prometteuse tentative de Greene et Cohen échoue à soutenir une thèse réformiste concernant la responsabilité criminelle.

4. Conclusion

Dans le présent texte, je me proposais de répondre à la question suivante: la compréhension du comportement humain que peut offrir les neurosciences menace-t-elle la responsabilité, au sens où elle est définie par le droit criminel? J'ai commencé par présenter ce qui constituerait un bon argument permettant de falsifier la catégorie juridique de responsabilité criminelle. Un tel argument doit montrer que l'ensemble des êtres capables d'exercer leur raison pratique est vide – ou presque vide – et donc que la catégorie ne peut jamais – ou pratiquement jamais – s'appliquer. Cela est cependant fort improbable. Malgré tout, j'ai montré qu'en changeant l'angle d'approche de manière à tenir compte des intuitions morales des individus dans l'application de la catégorie, Greene et Cohen peuvent formuler un argument valide qui falsifierait la catégorie simplement sur la base du déterminisme causal impliqué par les sciences de la nature, et s'appliquant explicitement au comportement humain dans le cas des neurosciences. Finalement, l'évaluation de l'argument dans sa substance m'a permis de montrer que, bien que valide, il n'est pas fondé, car il n'est pas établi que les individus ont bel et bien l'intuition morale nécessaire, soit qu'un individu n'est responsable d'une action que s'il en est la cause ultime.

Cela ne veut pas dire qu'aucune thèse réformiste ne peut être justifiée. Seulement, des arguments supplémentaires – ou une stratégie différente – sont nécessaires. Aussi, il est possible que la catégorie de responsabilité ait malgré tout à être complétée ou modifiée. Cependant, je crois qu'on aurait plus de chance de justifier une modification sur une base directement normative, par exemple pour mieux réaliser un idéal de liberté politique. Dans tous les cas, il faut faire bien attention avant d'écarter des catégories normatives aussi fortes au sein d'une société; elles aussi méritent un procès juste et équitable, surtout lorsqu'elles sont considérées passibles de la peine de mort³⁰.

1. À ce sujet, voir notamment les essais de Brian Leiter, «Naturalism and Naturalized Jurisprudence» dans B. Bix (dir.), *Analyzing Law: New Essays in Legal Theory*, Oxford, Clarendon Press, 1998 et «Naturalizing Jurisprudence: Three Approaches» dans J. Shook et P. Kurtz (dir.), *The Future of Naturalism*, Amherst, Prometheus Books, 2009.
2. Joshua Greene et Jonathan Cohen, «For the law, neuroscience changes nothing and everything» dans *Philosophical Transactions of the Royal Society B*, no. 359 (2004), pp. 1775-1785.
3. Pour une discussion détaillée des positions conservatrices et réformistes concernant l'impact des neurosciences sur le droit, voir Nicole Vincent, «Legal responsibility adjudication and the normative authority of the mind sciences» dans *Philosophical Explorations*, vol. 14, no. 3 (septembre 2011), pp. 315-331.
4. Voir, notamment, Michael Gazzaniga, *The Ethical Brain*, New York, Dana Press, 2005.
5. H. L. A. Hart, *Le concept de droit*, Bruxelles, Facultés universitaires Saint-Louis, 1976.
6. Stephen J. Morse, «New Neuroscience, Old Problems: Legal Implications of Brain Science» dans Walter Glannon (dir.), *Defining Right and Wrong in Brain Science. Essentials Readings in Neuroethics*, New York, Dana Press, 2007, p. 196. Voir aussi Hart, *op. cit.*, p. 115.
7. La prémisse pourrait aussi être «l'extension de P doit être suffisamment pleine», car une catégorie non vide mais à l'extension trop restreinte pourrait, dans certains cas, nuire tout autant à la visée pratique du droit.
8. H. L. A. Hart, *Punishment and Responsibility*, Oxford, Oxford University Press, 2008, pp. 181-182 et 201.
9. Pour plus de détails sur ces critères, voir la jurisprudence canadienne en la matière, notamment : R. c. Parks, [1992] 2 R.C.S. 871 (défense sur la base du somnambulisme); R. c. Stone, [1999] 2 R.C.S. 290 (défense d'automatisme avec aliénation mentale); R. c. Sault Ste. Marie, [1978] 2 R.C.S. 1299 (justification de la notion de *mens rea*); R. c. Hundal, [1993] 1 R.C.S. 867 (appréciation objective de la faute selon le critère de la personne raisonnable).
10. *Code criminel*, article 16. (1), L.R.C., 1985, ch. C-46.
11. R. c. Morin, [2012] QCCS 4722.
12. Cette idée est exprimée à l'article 17 du *Code criminel* du Canada.
13. Il est à noter qu'en droit criminel, à moins d'une disposition contraire, toute personne est préjugée remplir ces exigences et le fardeau incombe donc à l'accusé de démontrer qu'il ne doit pas être tenu responsable, s'il y a lieu.

14. Stephen J. Morse, «Brain Overclaim Syndrom and Criminal Responsibility: A Diagnostic Note» dans *Ohio State Journal of Criminal Law*, no. 3 (2006), p. 405.
15. Paul Churchland, «Eliminative Materialism» dans *The Journal of Philosophy*, vol. 78, no. 2 (1981), pp. 67-90.
16. *Ibid.*, p. 84.
17. Une description neuroscientifique est beaucoup plus complexe qu'une description dans les termes de la psychologie naïve. Il n'est pas clair que la complexité des neurosciences soit un gain pour les descriptions qui sont pertinentes pour la pratique du droit en comparaison, par exemple, à la pratique de la médecine ou de la psychiatrie.
18. Joshua Greene et Jonathan Cohen, *loc. cit.*
19. *Ibid.*, p. 1778.
20. Il pourrait certainement être contesté que les neurosciences rendent bel et bien manifeste un déterminisme causal fort. Je me permets cependant de leur concéder ce point pour les fins de l'argument.
21. Le «compatibilisme» est la thèse selon laquelle le libre arbitre et la responsabilité morale sont compatibles avec le déterminisme causal impliqué par les sciences de la nature.
22. Il y a en fait deux intuitions en jeu: (1) $R \rightarrow LA$, qui dit que le libre arbitre est une condition nécessaire de la responsabilité; et (2) $LA \rightarrow CU$, qui dit que le libre arbitre implique que l'agent soit la cause ultime de son action. Les deux, conjointement, forment l'intuition qu'une personne n'est responsable d'une action que si elle en est la cause ultime. Le droit refuse (1) et demeure neutre quant à (2). Voir aussi Stephen J. Morse, «The Non-Problem of Free Will in Forensic Psychiatry and Psychology» dans *Behavioral Sciences and the Law*, no. 25 (2007), pp. 203-220.
23. Greene et Cohen, *op. cit.*, pp. 1780-1781.
24. Le but ici est uniquement de montrer que la conclusion de Greene et Cohen ne découle pas nécessairement des prémisses. Il ne s'agit pas de démontrer l'interprétation alternative.
25. Katrina L. Sifferd, «Nanotechnology and the Attribution of Responsibility» dans *Neurotechnology Law and Business*, vol. 5, no. 2 (2008), pp. 177-190.
26. Greene et Cohen, *loc. cit.*, pp. 1781-1783.
27. *Ibid.*, p. 1782.
28. *Id.*
29. Pour une étude identifiant justement les intuitions naïves concernant le libre arbitre dans les termes de la médiatisation par des états mentaux

plutôt que dans les termes de CU, voir Dylan Murray et Eddy Nahmias, «Explaining Away Incompatibilist Intuitions» dans *Philosophy and Phenomenological Research*, 2012, à paraître [pré-publié en ligne].

30. Je remercie Jocelyn Maclure ainsi que les réviseurs du comité de rédaction de *Phares* pour leurs généreux commentaires sur des versions antérieures de cet article. Toute erreur ou imprécision n'est cependant attribuable qu'à l'état de mon propre cerveau.