Article

« Tirer la responsabilité au clair : le cas des attitudes implicites et le révisionnisme »

Luc Faucher

Les ateliers de l'éthique / The Ethics Forum, vol. 7, n° 1, 2012, p. 179-212.

TIRER LA RESPONSABILITÉ AU CLAIR : LE CAS DES ATTITUDES IMPLICITES ET LE RÉVISIONNISME

LUC FAUCHER
UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RÉSUMÉ

Dans cet article, je considère l'influence possible des recherches récentes sur les attitudes en psychologie sociale, principalement dans le paradigme des théories des processus duaux [dual process theories], sur notre compréhension de la responsabilité. La thèse que je soutiens est que certaines révisions à notre façon de comprendre la responsabilité et nos pratiques d'attribution de la responsabilité pourraient être justifiées par ces travaux. Avant de présenter les révisions que j'introduis, je décris les grandes lignes du paradigme que j'utiliserai, soit celui des théories processus duaux tel qu'appliqué aux attitudes. Puis, m'inspirant de Vargas (2004, 2005), je présente les différentes formes que peuvent prendre le révisionnisme. Parce que ces révisions s'appliquent à des notions qui sont utilisées à la fois par le commun des mortels et par les philosophes (qui tentent soit de les reconstruire rationnellement, soit de les modifier), je présente ce que l'on présume que pense chacun des groupes sur la question. Finalement, je présente trois révisions, plutôt « locales », que ces travaux pourraient inspirer.

ABSTRACT

In this paper, I want to consider the possible influence of recent research in social psychology about attitudes — more precisely in the dual process paradigm — on our comprehension of responsibility. The thesis I want to defend is that this body of work could inspire a certain number of revisions to our way of understanding responsibility. Before introducing these revisions, I describe, in broad strokes, the empirical paradigm of dual process as applied to attitudes. Then, in order to be clear about the kind of revisions that I will propose, I present Vargas' taxonomy of revisionism (2004, 2005). Because revisions are applied to notions that are used both by folks and philosophers (that present their theories as reconstructions of common sense), I present what each group thinks of responsibility. Finally, I present three revisions — mostly local revisions — that theories of dual attitudes could inspire.

Quiconque est un tant soit peu familier avec la littérature sur la responsabilité (ainsi que sur les attitudes et les pratiques qui lui sont associées) sait que ce concept a été maintes fois remis en question au cours de l'histoire de la philosophie. Que ce soit parce qu'ils acceptaient la vérité du déterminisme (pour des raisons théologiques, philosophiques ou scientifiques) et pensaient que le déterminisme était incompatible avec notre concept de responsabilité (ou du moins, certaines versions de celui-ci, Hobbes, 1645/1977 ; Spinoza, 1677/1994 ; Pereboom, 2001 et Smilansky, 2002 pour des versions contemporaines de ces arguments), ou parce qu'ils croyaient que l'idée que nous soyons à l'origine de nos actions dans un sens satisfaisant pour l'attribution de la responsabilité est une condition tout simplement impossible à réaliser (Strawson, 1994), les philosophes ont proposé soit d'abandonner le concept (et les attitudes et pratiques qui lui sont associées), soit de réviser de façon assez substantielle notre façon de penser la responsabilité. Récemment, des philosophes (Nahmias, 2006 ; Nelkin, 2005) ont noté que des « menaces » concernant l'idée que nous sommes responsables pourraient bien venir d'ailleurs, principalement de la psychologie (entendue au sens large, incluant les neurosciences cognitives). En effet, nombre de travaux (que ce soit ceux du neurologue Benjamin Libet et de ses collègues (1983) sur l'initiation de l'action, ou ceux de psychologues comme ceux de Wegner sur l'identification de la source de l'action et sur le contrôle de soi (2002), ou encore ceux de Ross et Nisbett (1991), sur l'importance des déterminants « situationnels » de l'action) ont été interprété par les philosophes (et parfois les psychologues eux-mêmes) comme suggérant que nous ne sommes pas responsables de nos actions, et qu'au mieux, nous n'avons jamais que l'illusion de l'être. C'est à un sous-ensemble de cette littérature que je vais m'intéresser aujourd'hui. J'aimerais considérer l'influence possible des recherches récentes sur les attitudes en psychologie sociale, principalement dans le paradigme des théories des processus duaux (*dual process theories*), sur notre compréhension de la responsabilité (et sur nos attributions de responsabilité). La thèse que j'aimerais soutenir est que certaines révisions à notre façon de comprendre la responsabilité et nos pratiques d'attribution de la responsabilité pourraient être justifiées par ces travaux. Disons d'entrée de jeu que je ne tenterai pas d'être exhaustif et de présenter l'ensemble des révisions possibles, mais uniquement de suggérer certaines pistes de réflexion concernant notre concept ou notre usage du concept (ainsi que les pratiques reliées à celui-ci) qu'offrent ces travaux. Je n'entends pas non plus offrir un ensemble unifié de révisions. Je tente plutôt d'illustrer le potentiel de ces nouvelles recherches pour la réflexion philosophique sur la responsabilité.

Avant de présenter les révisions que je propose, j'aimerais d'abord introduire rapidement le paradigme que j'utiliserai, soit celui des théories des processus duaux tel qu'appliqué aux attitudes (section 1). Puis, m'inspirant de Vargas (2004, 2005), je présenterai les différentes formes que peut prendre le révisionnisme (section 2). Les distinctions que cet auteur introduit au sujet du révisionnisme sont importantes pour mon propos. Comme on le verra, Vargas distingue une forme de révisionnisme global de formes plus parcellaires. Ce sont ces dernières qui m'intéresseront. En bref, je ne m'attaquerai pas à la question de sa-

voir si, à la lumière des travaux en psychologie sociale, notre concept de responsabilité devrait être abandonné ou révisé *in toto*, mais je proposerai plutôt certaines révisions, plutôt « locales », que ces travaux pourraient inspirer (section 4). Parce que ces révisions s'appliquent à des notions qui sont utilisées à la fois par le commun des mortels et par les philosophes (qui tentent soit de les reconstruire rationnellement, soit de les modifier), je présenterai avant ce que l'on présume que pense chacun des groupes sur la question (section 3).

1. LES THÉORIES DES ATTITUDES DUELLES

Connues des philosophes principalement parce qu'elles ont été invoquées par les psychologues du raisonnement comme Kahneman (2002) ou Evans (2003 ; voir Frankish et Evans, 2009 pour une histoire et un survol de ces théories), mais plus récemment par les psychologues travaillant sur les capacités morales (Haidt, 2001 ; Sunstein, 2005), les théories des processus duaux de l'esprit ont été proposées dans divers champs de la psychologie, dont entre autres, en psychologie sociale. C'est aux propositions des psychologues sociaux que je m'intéresserai dans ce qui suit, plus précisément au développement des théories des attitudes duelles[1].

Pour les partisans des théories des processus duaux, l'univers des processus mentaux peut être divisé en deux grandes classes générales : ceux qui opèrent automatiquement et ceux qui sont sous le contrôle volontaire. Plus précisément, comme le note Frankish (2010), les partisans d'une théorie des processus duaux, que ce soit en psychologie du raisonnement ou en psychologie sociale, affirment

> [qu'il] existe deux modes de traitement disponibles pour une tâche cognitive donnée, ceux-ci employant des procédures différentes et pouvant mener à des résultats conflictuels. Un premier type de processus (type 1) est rapide, automatique et non conscient, alors que le second (type 2) est plutôt lent, contrôlable et conscient (914 ; toutes les traductions de l'anglais vers le français dans ce texte sont de l'auteur).

Comme nous l'avons noté plus haut, des théories des processus duaux ont été également défendues en psychologie sociale (voir Smith et Collins, 2009 pour un survol des différents domaines où elles ont été proposées), notamment au sujet de ce que l'on nomme les « attitudes ». Il importe donc d'abord de préciser ce que l'on entend par « attitude » avant de décrire ce qu'affirment les théories des attitudes duelles[2].

En psychologie, le terme « attitude » fait référence à un ensemble de pensées, de sentiments et mêmes de tendances à l'action, à valence positive ou négative, que l'on entretient vis-à-vis un objet cible (une personne, un groupe social, un parti politique, etc.). Le terme « attitude » regroupe donc un ensemble d'états hétérogènes, soit des états cognitifs (des stéréotypes, des associations d'idées), des états affectifs (émotions, dispositions émotionnelles) et des tendances comportementales (approche, évitement)[3]. Comme leurs analogues dans les autres do-

maines de la psychologie, les théories des attitudes duelles postulent que des attitudes de types différents (la typologie étant établie en fonction de l'accessibilité de ces attitudes) peuvent coexister et affecter le comportement et la cognition[4]. Et, comme dans les autres domaines, une grande partie du travail des psychologues consiste à décrire l'étendue de la dissociation entre les deux types d'attitudes et l'effet de cette dissociation sur le comportement.

Les théories des attitudes duelles ont été proposées parce qu'elles permettaient de rendre compte d'une dissociation entre ce que les agents rapportent au sujet de leur attitude vis-à-vis d'un stimulus, et leur comportement vis-à-vis de ce même stimulus. Selon ces théories, nos attitudes face aux gens considérés comme appartenant à des races, genres sexuels, ou groupes sociaux différents (par exemple, les personnes âgées, les handicapés ou les gens obèses) peuvent être de deux types. Le premier type est composé des « attitudes explicites », celles que nous reconnaissons avoir vis-à-vis des gens appartenant à ces groupes. En ces temps où il existe un consensus moral à l'égard du racisme et de la discrimination, ces attitudes sont souvent des attitudes égalitaristes ou allant à l'encontre des stéréotypes. Ainsi, lorsqu'on nous le demande (ou lorsque nous sommes soumis à un test sondant notre opinion), nous pouvons rapporter que nous n'entretenons pas d'attitudes négatives vis-à-vis des afro-américains, des obèses ou des personnes âgées. Le second type d'attitudes est composé des « attitudes implicites », c'est-à-dire d'attitudes que les gens ne sont pas conscients d'avoir, et qui sont activées pour ainsi dire à leur insu[5]. Au cours de la dernière décennie, les outils pour étudier ces dernières se sont affinés[6] et ils ont permis de montrer que les gens pouvaient rapporter n'entretenir explicitement aucune attitude négative vis-à-vis d'un groupe, mais, en même temps, afficher implicitement des attitudes négatives vis-à-vis du même groupe (par exemple associer implicitement les individus appartenant à un groupe à la violence ou à la paresse). Une théorie des attitudes duelles fera donc l'hypothèse que nous pouvons avoir des évaluations différentes (parfois même contradictoires) et différentes au niveau de l'accessibilité, d'un objet, une personne ou un groupe : une attitude implicite et une attitude explicite.

Selon certains psychologues, le seul fait de percevoir un membre d'un groupe à l'égard duquel nous avons des attitudes implicites est suffisant pour activer ces dernières. Par exemple, John Bargh, qui est un des tenants de cette théorie des attitudes duelles, écrit que « la simple perception de caractéristiques facilement discernable d'un groupe [… est] suffisante pour […] causer l'activation du stéréotype associé au groupe » (1999, 363 ; traduction libre). Plus récemment, il écrivait que dans certains cas, « la simple présence de certains évènements ou de certaines personnes activerait automatiquement nos représentations de ceux-ci, et de façon concomitante, toute l'information interne […] emmagasinée qui est pertinente pour leur répondre » (Bargh and Morsella, 2008, 76 ; traduction libre).

Pourquoi devrait-on s'en faire au sujet de ces attitudes activées automatiquement et à nos dépens ? L'existence de ces attitudes n'aurait pas beaucoup d'importance pour nous si elles prédisaient de façon peu fiable notre comportement.

Malheureusement, comme le note encore Bargh, la recherche en psychologie sociale « a commencé à montrer que des comportements complexes sont aussi informés et guidés automatiquement par la connaissance qui est activée incidemment pendant la perception » (Ferguson et Bargh, 2004, 33 ; traduction libre). Pour illustrer le problème que pose cette situation, j'aimerais emprunter un exemple particulièrement dramatique de l'effet sur le comportement des attitudes raciales implicites que mes collègues Edouard Machery, Dan Kelly et moi-même avons utilisé dans quelques articles récents (Faucher et Machery, 2009 ; Machery, Faucher, Kelly, 2010 ; Kelly, Faucher, Machery, 2010) : le biais lié aux armes (*weapon bias*).

Selon Keith Payne (pour un résumé de ses travaux sur la question, voir son 2006) à qui l'on doit la découverte de ce biais (et le paradigme pour l'étudier), les individus dont le niveau d'attitudes implicites négatives est plus élevé à l'égard des noirs (voir également les travaux récents au sujet des arabes et de ce que l'on a nommé « l'effet turban », Unkelbach et al., 2008) qu'à l'égard des autres tendent à manifester un biais lié aux armes plus important. En quoi consiste ce biais ? En gros, il consiste dans un patron de réponses suivant à certaines conditions. Dans des conditions où ils n'ont pas beaucoup de temps[7], on demande à des sujets[8] d'identifier des porteurs d'arme et de tirer sur eux, mais également d'éviter de tirer sur ceux qui sont désarmés et qui n'ont par exemple qu'un contenant de boisson gazeuse ou un téléphone à la main. Dans ces conditions, ceux chez qui on a mesuré un niveau d'attitudes implicites négatives plus élevé que chez les autres détectent plus rapidement une arme et tirent plus rapidement lorsqu'ils voient un Afro-Américain que lorsqu'ils voient un caucasien. Plus encore, ils tendent à faire plus souvent l'erreur de voir une arme dans les mains d'un Afro-Américain qui n'a qu'un contenant de boisson gazeuse ou un téléphone à la main, et à ne pas identifier l'arme dans la main d'un caucasien et à penser qu'il a un contenant de boisson gazeuse ou un téléphone.

L'effet des attitudes implicites sur le comportement ne se limite pas à ce cas. Les psychologues sociaux ont effectivement montré que dans certains cas, elles peuvent influencer les décisions d'embauche, les décisions médicales, les décisions politiques, la coopération et la qualité des interactions entre individus appartenant à des groupes raciaux différents, le comportement d'assistance à une victime dans le besoin (pour une revue de ces effets, voir Pearson et al. 2009). Ces attitudes engendrent donc, dans un ensemble de domaines variés, des comportements moralement problématiques.

Une question importante devient de savoir dans quelle mesure ces attitudes implicites ont une emprise sur notre vie. Alors que Payne semble vouloir restreindre leurs influences à des conditions bien particulières (celles où nos ressources cognitives et attentionnelles sont affaiblies), John Bargh croit plutôt qu'« [u]ne grande partie de la vie quotidienne d'une personne est déterminée non par ses intentions conscientes et ses choix délibérés, mais par des processus mentaux qui sont mis en action par des caractéristiques de l'environnement opérant à l'extérieur de la conscience et du contrôle conscient » (Bargh et Chartrand, 1999, 462).

Comme on peut l'inférer de la dernière partie de la citation, Bargh est particulièrement sceptique quant à la possibilité de contrôler ces attitudes. D'une part, les conditions nécessaires au contrôle (par exemple, la conscience que les stéréotypes opèrent, la motivation d'agir sur ceux-ci ainsi que la connaissance des moyens de les neutraliser) sont rarement réunies. D'autre part, même lorsque ces conditions sont réunies, « les jugements et les comportements basés sur les stéréotypes peuvent néanmoins se produire » (Bargh, 1999, 371). Selon lui, « à ce jour les données concernant les possibilités réalistes que les gens contrôlent les stéréotypes activés automatiquement penchent largement du côté négatif » (idem, 378).

Si la situation est bien telle que la décrit Bargh, on pourrait penser que le domaine des actions dont nous sommes responsables devient soudainement extrêmement étroit, beaucoup plus que ce que la plupart d'entre nous ne se l'imaginent. En effet, si une grande partie de nos actions échappe à notre contrôle et se fait à notre insu, on pourrait être tenté de dire que nous ne sommes pas les auteurs de ces actions, et donc que nous ne sommes pas responsables de celles-ci[9]. C'est ce que Bargh nomme « l'implication terriblement déprimante » (« *the tremendously depressing implication* ») de ces travaux (1999, 363). C'est pour cette raison que certains considèrent que les travaux des partisans des théories des attitudes duelles représentent une menace pour la responsabilité. La menace n'est, à proprement parler, pas une menace au concept de responsabilité lui-même, mais plutôt à son extension présumée. Si ce qu'avancent ces chercheurs se révèle exact, nous serions sous l'emprise d'une illusion quant à la véritable étendue de notre responsabilité[10].

Mais ne pourrions-nous pas tenir pour avérés les résultats de ces travaux et ne pas accepter l'implication déprimante de Bargh ? En d'autres mots, ne pourrions-nous pas tout de même être considérés responsables des actions provoquées par nos attitudes implicites ? Il semble à première vue que cela nous forcerait à aller à l'encontre de certaines de nos intuitions concernant la responsabilité. Mais est-ce bien le cas ? Notre concept ne permet-il pas que nous soyons responsables d'actions que nous ne faisons pas volontairement ? Il semble que, pour répondre à cette question, il faille éclaircir ce concept qui guide nos jugements. C'est ce que j'aimerais faire dans ce qui suit. Mais, pour des raisons qui seront bientôt évidentes, j'aimerais inscrire ma démarche dans le cadre du révisionnisme.

2. LE RÉVISIONNISME

Puisque mon intention est d'inscrire ma démarche dans un cadre révisionniste, il importe de préciser ce que j'entends par cette notion. Ma compréhension de la question est tributaire dans une large mesure du cadre proposé par Manuel Vargas, dont je vais présenter dans ce qui suit les grands traits (tels que développés principalement dans son « The Revisionist's Guide to Responsibility » 2005 ; voir également Vargas 2004).

Lorsque les philosophes s'intéressent à la question de la responsabilité, nous dit Vargas, ils s'intéressent habituellement à un ou à plusieurs des éléments suivants : 1) les dispositions ou les attitudes associées à la responsabilité (par exemple, la colère et l'indignation, ou encore la culpabilité) ; 2) aux pratiques associées à la responsabilité (par exemple, le blâme et la punition) ; et finalement : 3) aux croyances et aux concepts concernant la responsabilité (ce que l'on pourrait qualifier de « concepts populaires » [*folk concepts*] de la responsabilité).

Au sujet de ces éléments, les philosophes se posent essentiellement trois types de questions. D'abord des questions d'ordre *métaphysique* du type « Quelle est la nature de la responsabilité ? » (De quoi dépend-elle ou à quoi est-elle reliée ? Par exemple, dépend-elle d'une relation particulière de l'agent à son action ?). Ensuite, des questions plutôt *descriptives* du type « Que pense-t-on à propos de la responsabilité[11] ? » ou « À quelle condition pense-t-on que l'on peut appliquer le concept de responsabilité ? » (Quel est notre concept de responsabilité ?). Enfin des questions *normatives* comme « Comment devrions-nous penser la responsabilité ? » ou « Quel devrait être notre concept de responsabilité ? ».

Pour un non-révisionniste, il existe en quelque sorte une « harmonie préétablie » (Vargas, 2005, 411) entre ce que nous pensons de la responsabilité et ce que nous devrions en penser, compte tenu de ce que nous savons de la responsabilité elle-même ou du monde.

Le révisionniste ne partage pas cette croyance et suppose pour sa part une différence entre les deux derniers projets (soit le projet descriptif et le projet normatif) telle que, ce que prescrit le second diffère de ce que décrit le premier. Ainsi, « toute théorie qui soutient que notre façon de penser est telle, mais devrait être autre, sera une instance paradigmatique de révisionnisme[12] » (2005, 404) ou, plus précisément, un cas paradigmatique de révisionnisme est une « théorie qui prescrit une révision, relative à une approche diagnostique [descriptive] de notre façon de penser coutumière, dans nos pratiques, attitudes ou croyances par rapport à la responsabilité et nos conceptions caractéristiques de celle-ci » (421).

Selon Vargas, la force du révisionnisme peut varier de telle façon qu'il est possible de parler de révisionnisme faible, modéré et fort. Un révisionnisme sera dit *faible* si ce ne sont pas nos concepts, pratiques et attitudes qui demandent à être révisés, mais plutôt notre compréhension de ce en quoi consiste ceux-ci (à la suite, par exemple, de la découverte que nous sommes mépris sur ce que nous pensions être nos propres concepts, pratiques ou attitudes ou ceux que partageait notre communauté). Par contraste, un révisionnisme sera dit *fort* si ce qu'il exige est l'élimination pure et simple de nos concepts, pratiques ou attitudes parce qu'ils posent, par exemple, des demandes métaphysiques ou psychologiques qu'il est impossible de satisfaire. Un révisionnisme *modéré*, lorsqu'appliqué par exemple au concept de responsabilité, « est l'idée que notre concept populaire de responsabilité sera jugé inadéquat jusqu'à ce qu'il soit modifié de

quelque façon […]. [Cette dernière forme de révisionnisme] revient à une forme d'émondage [*pruning*] de cet élément » (Vargas, 2005, 409). Il pourrait par exemple consister dans le fait de se débarrasser d'un élément jugé superflu de notre concept, comme celui d'auto-détermination ou d'autonomie.

Non seulement la force, mais aussi l'étendue du révisionnisme peut varier. Par exemple, celui que Vargas nomme « un révisionniste sophistiqué » (412) n'exigera pas que la révision du concept de responsabilité implique qu'une telle révision doive être appliquée en même temps aux pratiques et aux attitudes (c'est ce que j'appelais en introduction une révision *in toto* que je distinguais d'un « révisionnisme parcellaire » qui est celui proposé par le révisionniste sophistiqué). Le révisionniste sophistiqué s'oppose donc au révisionniste « en bloc » qui propose de réviser (le plus souvent de rejeter) à la fois le concept, les pratiques et les attitudes. Par ailleurs, le révisionniste sophistiqué pourra avoir au sujet de chacun des objets une attitude différente : par exemple, il pourrait vouloir se débarrasser du concept de responsabilité, conserver les attitudes et certaines pratiques, et se débarrasser de certaines autres (par exemple, se débarrasser de la punition, mais conserver le blâme). Le révisionniste sophistiqué peut également proposer des changements variés à l'intérieur même d'une catégorie. Par exemple, il pourrait exiger que l'on se débarrasse du lien entre le concept de responsabilité et l'existence de phénomène *causa sui* (l'idée que l'on puisse se déterminer soi-même), mais conserver l'idée que certaines conditions épistémiques, ou celles liées à la conscience, fassent partie du concept.

Maintenant, quelles sont les contraintes que doit respecter celui qui propose une révision modérée (ou même faible) ? Selon Vargas encore, il existe deux contraintes : la révision doit être *plausible* et elle doit être *justifiée normativement* (dans Vargas 2004, il les nomme respectivement la *contrainte de plausibilité naturaliste* et *l'adéquation normative*). En ce qui concerne la plausibilité, il écrit : « une révision requérant quelque chose qui n'est pas psychologiquement possible ou qui n'est pas plausible socialement a de fortes chances d'être rejetée ou devrait l'être » (413). Quant à la contrainte de justification normative, la tâche consiste à montrer que la révision peut être justifiée sur la base de raisons indépendantes de faits métaphysiques qui sont considérés comme des conditions nécessaires à l'application de notre concept de responsabilité (par exemple, l'existence d'une volonté libre), particulièrement si la révision est induite par l'absence de faits exigés par le concept de responsabilité. Par exemple, Nichols (2007) accepte la vérité du déterminisme et croit que cela devrait le conduire à considérer que nous ne sommes pas responsables (parce que cela serait incompatible avec l'existence d'une volonté libre), mais il propose des raisons évolutionnistes et utilitaristes de conserver certaines de nos attitudes réactives (certaines formes de punition des transgressions de normes sociales sont, selon lui, un élément essentiel du fonctionnement des groupes humains). On pourra également faire référence à d'autres aspects des théories éthiques (par exemple, des considérations sur la justice ou l'équité) comme justification d'une révision[13]. Notons, et c'est important pour ce qui suit, que pour un révisionniste faible, la justification d'une révision viendra souvent d'études empiriques

(celles-ci ayant, peut-on supposer, un plus haut degré d'autorité épistémique que nos propres croyances au sujet du concept) : « Dans la mesure où de telles études empiriques altèrent notre compréhension coutumière de nos pratiques ou de nos attitudes caractéristiques de la responsabilité, de telles théories peuvent compter comme des théories faiblement révisionnistes de la responsabilité » (Vargas, 2005, 414). En gros, les études empiriques pourraient nous montrer que les gens ont un concept de responsabilité (ou des attitudes ou pratiques) différents de celui que nous leur imputons.

Ce que j'aimerais proposer dans ce qui suit, c'est que les travaux sur les attitudes implicites contiennent les germes de révisions de notre façon de comprendre divers aspects de la responsabilité. Comme on le verra, ces révisions ne sont pas des révisions fortes, mais plutôt des révisions faibles ou modérées. Mon hésitation quant à la force de ces révisions vient entre autres du fait que nous n'avons pas une vision très claire de ce que nous pensons communément de la responsabilité. Il est également possible que nos intuitions à ce sujet ne soient pas bien stables.

3. CE QUE PENSENT LE SENS COMMUN ET LES PHILOSOPHES

Revenons donc à la question qui nous occupe, celle de la responsabilité pour les actions induites par nos attitudes implicites. Le psychologue Keith Payne la résume bien lorsqu'il écrit :

> Le problème quant à ces décisions prises en une fraction de seconde est qu'elles semblent se faire d'elles-mêmes […]. Dois-je considérer ces décisions comme *mes* décisions si elles diffèrent de mes intentions ? Qui est responsable ? (2006, 287).

Demandons-nous d'abord ce que le commun des mortels puis ce que les philosophes pensent de la question. C'est sur la base de ce que les premiers pensent et que les seconds leur imputent que, rappelons-le, nous pouvons proposer des révisions[14].

3.1 LE SENS COMMUN

Que pense le commun des mortels de la responsabilité concernant les attitudes implicites ? Une façon de le savoir est de poser la question directement aux gens. C'est ce que font Cameron, Payne et Knobe (2010) dans leur article « Do Theories of Implicit Race Bias Change Moral Judgments ? ». Cameron et ses collègues exploitent dans ce texte un désaccord dans la littérature psychologique au sujet des attitudes implicites. Il faut en effet savoir qu'il existe deux versions de la théorie des attitudes implicites. Selon une théorie, celle qui est la plus répandue, l'effet des attitudes implicites est automatique et inconscient. Selon une autre théorie, les attitudes implicites sont seulement automatiques, les sujets sachant habituellement qu'ils les possèdent (leur problème étant généralement plutôt qu'ils pensent que ces biais n'ont pas d'effet sur le comportement et que pour cette raison, ils ne prennent pas la peine de les contrôler).

Dans une de leurs expériences, les trois auteurs ont proposé une des vignettes suivantes à des sujets :

1) Donald doit attribuer une promotion à un de ses employés. Il pense que les gens devraient être traités de façon équitable, quelle que soit leur race. Malgré cela, Donald a une aversion inconsciente pour les Afro-Américains. Il ne sait pas qu'il a une telle aversion, et s'il le savait, il la désapprouverait parce qu'il croit sincèrement à l'égalité. Son aversion inconsciente a une influence sur son comportement, dont il ne se doute pas. Lorsqu'il doit prendre une décision concernant la promotion d'un de ses employés, il tente de le faire entièrement sur la base du mérite, mais il ne réussit pas toujours à prévenir l'influence de son aversion sur son jugement. En conséquence, il refuse la promotion de façon équitable à des Afro-Américains.
2) L'histoire est identique, mais, plutôt qu'avoir une aversion inconsciente, Donald a une sensation viscérale (un « *gut feeling* ») dont il est conscient, mais qu'il ne peut pas contrôler.
3) L'histoire est identique, mais, malgré ses bons sentiments, Donald refuse d'accorder une promotion de façon équitable à des Afro-Américains. On ne fait aucune mention des attitudes inconscientes.

Pour chacune des vignettes, les chercheurs ont demandé aux sujets d'indiquer, sur une échelle de 1 à 5, s'ils considéraient que Donald était responsable, qu'il devait être blâmé ou bien puni (1 signifiant qu'ils ne le considèrent pas du tout responsable, et 5 qu'ils le considèrent pleinement responsable). À la lumière des résultats qu'obtiennent Cameron et ses collègues, il apparaît clairement que, dans le cas où le comportement répréhensible est causé par une aversion automatique, mais consciente, les sujets jugent Donald comme étant presque aussi responsable que lorsqu'il agit sciemment. Ce n'est que lorsque l'aversion est automatique *et* inconsciente que leur jugement de responsabilité devient moins sévère.

Cameron et ses collaborateurs tirent la conclusion suivante de leurs travaux : « L'automaticité consciente et l'automaticité inconsciente ont […] des implications différentes dans l'assignation de responsabilité morale » (19). Ces expériences semblent donc révéler que c'est le fait d'être conscient ou non de l'attitude implicite qui fait la différence dans l'attribution de responsabilité.

3.2 LES PHILOSOPHES

Pour l'instant, il y a peu de philosophes qui se sont préoccupés de la question de la responsabilité pour les actions produites par les attitudes implicites. Parmi ceux-ci, deux semblent reprendre dans les grandes lignes le jugement du sens commun, soit Levy (2008) et Wigley (2007). Les deux philosophes rejettent la responsabilité essentiellement pour la même raison. Au demeurant, ils pourraient tous deux être classés comme des partisans de « la théorie volitionniste » de la responsabilité (comme l'a nommé Smith, 2005) selon laquelle le choix préalable, la décision ou le contrôle volontaire est l'élément crucial pour l'attribution de la responsabilité.

Selon Levy, il résulte du fait qu'on ne puisse consentir consciemment à l'action provoquée automatiquement par l'attitude implicite que nous ne pouvons être tenu responsable de celle-ci. Comme il l'écrit :

> Lorsque [les attitudes] sont produites automatiquement et ne sont ni le produit direct ni le produit indirect de l'approbation consciente, nous ne sommes pas responsables [de nos comportements]. Nous n'avons pas choisi de répondre de la façon dont nous le faisons, et donc, nous n'exerçons pas de contrôle, direct ou indirect, sur nos actions (2008, 219-220).

Levy semble suggérer deux raisons pour lesquelles nous ne pouvons être tenus responsables de nos actions lorsque celles-ci sont causées par nos attitudes implicites. Ces deux raisons dépendent du fait qu'il suppose que les attitudes implicites sont inconscientes. La première raison consiste à soutenir que, puisque l'attitude implicite est inconsciente, nous n'avons pas pu lui donner notre approbation. L'idée que nous sommes responsables de ces actions uniquement si elles procèdent des volitions (ou des attitudes) que nous endossons ou que nous avons choisies après délibération est une idée qui est défendue dans la littérature, entre autres par Frankfurt (1971). Comme l'écrit Levy :

> Les actions sont profondément attribuables aux agents uniquement lorsque la conscience joue un rôle substantiel dans leur production […]. La délibération consciente reflète proprement la personne entière, y compris ses valeurs consciemment endossées (2008, 217 et 220).

La seconde raison est que, puisque nous ne sommes pas conscients de ces attitudes, nous ne pouvons pas exercer de contrôle sur les comportements qui en découlent[15]. Puisque nous ne pouvons contrôler ces comportements, nous ne pouvons en être tenus responsables. Cette seconde raison est en fait une reprise d'une des conditions (nécessaire, mais non suffisante) classiques de la responsabilité (telles qu'énoncées par Aristote dans le livre III de l'*Éthique à Nicomaque*), soit la condition de contrôle. Un agent est dit responsable d'une action si et seulement si il a la capacité de la poser ou non (s'il n'est pas contraint). Si, dans un sens, cette raison est la reprise de la condition classique, on pourrait dire qu'elle en est une version puisqu'elle pose que nous ne pouvons contrôler les actions dont nous ne sommes pas conscients, ce que ne semble pas exiger nécessairement le fait de pouvoir contrôler une action (comme nous le verrons plus loin). En effet, le simple fait d'être conscient de quelque chose ne nous assure pas que nous pouvons la contrôler, tout comme le fait que nous ne soyons pas conscient d'une chose n'empêche pas que notre comportement puisse être contrôlé. Le problème des attitudes implicites, à ce moment-là, viendrait donc plus de leur automaticité présumée que du fait de n'en être pas conscient.

C'est à cette question que s'attaque Wigley (2007) dans son article « Automaticity, Consciousness and Moral Responsability ». À première vue, ce dernier semble plutôt appartenir au groupe des philosophes qui soutiennent que nous sommes responsables des actions causées par nos attitudes implicites, et s'éloi-

gner du sens commun. En effet, Wigley écrit : « le fait que l'agent agissait "sans penser" lorsque le mal [*harm*] est survenu ne lui épargne pas une évaluation morale » (2007, 223).

Mais ce que Wigley soutient est que la présence de la conscience ou le contrôle n'est pas nécessaire au moment où l'agent commet l'action. Elle est cependant nécessaire soit avant, soit après l'exécution de l'action ou, comme il l'écrit :

> [u]ne certaine place [*elbowroom*] apparaît pour l'évaluation morale si nous portons notre attention sur le contrôle d'une personne quant à l'acquisition et la révision du comportement automatisé [*habituated behavior*] (2007, 211).

Ainsi, si une personne savait qu'en acquérant un stéréotype S elle risquait d'agir à son insu de façon moralement répréhensible dans le futur ou bien, si, s'apercevant qu'elle nourrit des stéréotypes à l'endroit d'un groupe, elle ne fait rien pour réviser ce stéréotype, alors elle est responsable. Comme l'écrit Wigley :

> La thèse principale que je soutiens est que nous exerçons un contrôle rationnel sur notre comportement automatique pour autant que nous puissions programmer (*preset*) ou réviser la façon dont il fonctionne (2007, 218).

Dans ce cadre, la réponse à certaines questions empiriques devient importante : exerçons-nous une forme de contrôle sur l'acquisition des stéréotypes ? Et si ce n'est pas le cas, pouvons-nous a moins les réviser (ou les contrôler) après qu'ils soient en place ?

Pour ce qui est de la première question, disons que l'affaire semble entendue (pour un compte rendu des travaux sur la question, voir Dunham et al., 2008, et Olson et Dunham, 2010). Nous acquérons certaines attitudes implicites très tôt dans le développement, vers 4 ou 5 ans, bien avant que nous soyons en âge d'avoir un avis éclairé sur celles-ci[16]. En ce qui concerne la seconde question, disons que, même si des psychologues comme Bargh sont assez pessimistes, tous ne partagent pas leur évaluation de la situation. Comme on le verra dans la dernière section de cet article, certains pensent qu'il est possible, sinon de se débarrasser de quelques-unes de nos attitudes négatives, du moins de les contrôler.

Reste un cas, celui où l'attitude responsable du comportement est acquise à l'insu de l'agent et impossible à réviser. Dans ce cas, affirme Wigley, « l'évaluation morale […] pourrait dépendre de la question de savoir s'il endosse son comportement automatique après réflexion ; c'est-à-dire s'il aurait agi ainsi même en l'absence de l'automaticité » (2007, 218). Cette condition est typiquement celle que proposent les théoriciens du « vrai moi » (comme les nomme Wolf, 1993) que sont Frankfurt et Watson, et celle que proposait Levy. Si Wigley accepte leur théorie, il propose cependant un amendement à celle-ci, celui qui consiste à dire qu'il est du devoir de l'agent, ou bien au moment où il acquiert un automatisme, ou bien au moment où il se rend compte de l'action de cet automatisme, de faire ce qui est en son pouvoir pour se prémunir contre ses effets

moralement négatifs. On pourrait ainsi dire que Wigley ajoute une condition aux théories du vrai moi : il n'est pas suffisant de rejeter, après réflexion, un état mental comme n'étant pas le sien ; il faut encore, si cela est possible, faire tout ce qui est en son pouvoir pour qu'il ne cause pas de tort moral.

3.3 PROJET DESCRIPTIF

Comme je l'ai mentionné à la section 2, la discussion révisionniste que j'entends proposer exige que l'on ait procédé à la description de nos croyances par rapport à la responsabilité et de celles des philosophes. Dans cette section, j'ai donc présenté une étude qui prétend établir ce que nous pensons à ce sujet. J'ai également présenté le point de vue de philosophes dont l'opinion converge avec ce que nous révèle cette étude de nos intuitions concernant nos attitudes implicites.

4. LES RÉVISIONS

Dans cette section, j'aimerais faire trois propositions à saveur révisionniste. Pour être clair, disons tout de suite que je n'entends pas proposer une révision globale et en profondeur de notre concept de responsabilité (ou de nos attitudes et pratiques associées), ni même présenter un ensemble unifié de révisions, mais plutôt suggérer quelques révisions que pourraient inspirer les travaux sur les attitudes implicites à nos conceptions du sens commun sur la responsabilité (et à certaines positions philosophiques qui se veulent en continuité avec les conceptions du sens commun). Est-ce que ces révisions en inspireront de plus profondes, plus fortes, mettront-elles les philosophes sur la voie de la reconstruction ou de l'élimination du concept de responsabilité ? Cela est possible et je ne veux pas répondre à cette question ici. Mon souhait est de convaincre ceux qui ne le seraient pas que la question des attitudes implicites demande une réflexion importante de la part des philosophes et de suggérer quelques voies de réflexion à ceux-ci. Dans cette optique, je proposerai les trois révisions suivantes :

1) La première concernant le rôle de la connaissance consciente dans l'attribution de la responsabilité ;
2) La deuxième concernant la théorie du vrai moi, telle que proposée par Frankfurt et Watson (mais également par Levy et Wigley dans le contexte de notre discussion) ;
3) La troisième concernant le concept de contrôle utilisé dans la discussion sur la responsabilité.

Encore une fois, les révisions que je proposerai seront des révisions faibles, c'est-à-dire des révisions à nos façons de comprendre nos propres concepts, pratiques ou attitudes. Dans certains cas, les révisions que je proposerai auront pour objets les méthodes par lesquelles nous établissons que certains critères jugés nécessaires pour la responsabilité sont satisfaits.

4.1. LA CONNAISSANCE CONSCIENTE

Comme nous l'avons remarqué plus haut, pour le sens commun, le fait d'être conscient des demandes morales d'une situation semble être une condition nécessaire de l'attribution de la responsabilité. Pour cette raison, le fait que les

agents ne sachent pas qu'ils possèdent des attitudes implicites et que ces dernières agissent à leur insu ferait en sorte que l'on ne reconnaît pas les agents comme responsables des actions qui découlent de ces attitudes. Levy et Wigley semblent bien capturer et expliciter les intuitions du sens commun lorsqu'ils affirment que la conscience de l'attitude est nécessaire à l'attribution de responsabilité.

Le problème est que les intuitions du sens commun ne sont peut-être pas univoques. En d'autres mots, il est possible que l'idée selon laquelle il faille être conscient pour être responsable ne soit pas partagée par tous. En effet, comme le remarquaient en note de bas de page Cameron et ses collègues (2010), s'il est vrai que c'est le cas pour les participants caucasiens de leurs expériences, les participants afro-américains semblent avoir un autre avis sur la question.

> [I]l y avait une interaction significative entre la condition de la théorie et la race du participant […]. Les caucasiens présentaient le même patron de résultats que l'échantillon complet, c'est-à-dire que la condition inconsciente provoquait un jugement de responsabilité moindre qu'avec la condition automatique. Cependant, les Afro-Américains présentaient seulement un effet marginal de la condition de la théorie […]. Cela était dû au fait que leurs attributions de responsabilité étaient plus élevées au regard de la condition automatique qu'au regard de la condition du sens commun et de la condition inconsciente (278).

Cette tendance d'un groupe qui est l'objet de discrimination à attribuer une responsabilité même pour les actions qui ne sont pas le produit d'une intention consciente semble être confirmée dans la littérature. Par exemple, la professeure de droit, Amy Wax écrit que « les blancs tendent à voir l'intention comme un élément essentiel du mal racial, pas les noirs… » (1993, 18 ; cité par Wellman, 2007, 48). Hardin et Banaji relèvent une différence dans les sensibilités qui pourraient bien être pertinentes pour nous, en citant les recherches de « Richeson and Shelton (2005) [lesquelles] ont montré que, dans les interactions face-à-face, les différences individuelles dans les préjugés implicites étaient plus apparentes pour les personnes de race noire que pour les personnes de race blanche, et plus apparentes quand les blancs interagissaient avec les noirs qu'avec d'autres blancs » (sous presse, p.11). Ainsi, pour les individus de race noire, non seulement la présence d'une intention consciente n'est pas nécessaire pour qu'une action soit moralement répréhensible et que l'on attribue la responsabilité de cette action à celui qui l'a commise, mais les blancs sont en quelque sorte aveugles à leurs propres préjugés ou attitudes implicites ainsi qu'à ceux de leurs semblables. Comme l'expliquent Pearson et ses collaborateurs :

> La perception des blancs quant à la façon dont ils se comportent ou sont perçus par les autres est basée plus sur leurs attitudes explicites et leurs comportements visibles […] que sur leurs attitudes implicites ou leurs comportements moins délibératifs. Par contraste, la perspective des partenaires d'interaction noirs dans ces interactions leur permet de porter attention à la fois aux com-

> portements spontanés (par ex. non verbaux) et délibératifs (par ex. verbaux) des blancs […] (2009, p.10).

Pronin (2006) nomme cette tendance générale à donner plus de poids aux données introspectivement accessibles dans l'évaluation de nos actions et à donner moins de poids aux données externes (comme le comportement) « la tache aveugle à l'égard des préjugés » (*bias blind spot*). Selon elle, « les gens croient en général qu'ils sont immunisés contre les préjugés à l'égard des groupes […]. Ils proclament qu'ils ne sont pas l'objet de ces préjugés, […], même dans des circonstances où ils ont manifesté ces préjugés » (2006, 38).

Ce que suggèrent ces travaux est que non seulement nous souffrons de cécité ou paralysie morale vis-à-vis de certaines situations (nous ne semblons pas voir les demandes morales que posent ces situations ou ne réussissons pas à agir en fonction de celles-ci à cause de nos attitudes implicites), mais nous souffrons également de cécité à l'égard de cette cécité ou paralysie morale (une espèce de déni de l'aveuglement ou de la paralysie). Nous ne reconnaîtrions pas les traces des préjugés dans notre comportement et celui des gens appartenant à notre groupe.

Pronin décrit une tendance *générale* à l'aveuglement, c'est-à-dire une tendance à préférer un certain type d'évidences lorsque vient le temps d'évaluer notre propre comportement et celui des gens de notre groupe. Ce que j'aimerais suggérer, c'est que le fait d'être l'objet de la discrimination fait peut-être en sorte de déplacer le regard de ce que disent les agents à ce qu'ils font (à accentuer, en quelque sorte, la tendance de ceux qui sont l'objet de la discrimination à utiliser les données externes plutôt que les données « internes »[17]). Il risque donc d'y avoir un genre d'asymétrie entre les points de vue sur la responsabilité d'un agent selon que l'on est celui qui commet l'injustice ou le tort moral (dans ce cas-ci, la discrimination), et celui qui le subit.

Cela me conduit à la possibilité d'une première révision. Celle-ci s'inspire d'une discussion récente de Knobe et Doris (2010) sur les théories de la responsabilité. Dans cette discussion, les auteurs arguent que les travaux récents en philosophie expérimentale démontrent que nos attributions de responsabilité morale dépendent et changent en fonction de certaines variables contextuelles. Le « *variantisme* », comme ils nomment cette idée, s'oppose à « *l'invariantisme* » qui aurait dominé les discussions philosophiques jusqu'ici et selon lequel les gens « devraient appliquer les mêmes critères dans tous leurs jugements de responsabilité morale … [le variantisme, pour sa part, suppose plutôt que] si on veut comprendre pourquoi les gens font les jugements qu'ils font, il ne sert à rien de chercher un seul ensemble de critères de base qui s'accordent avec tous leurs jugements ordinaires. Une approche plus prometteuse consiste à examiner comment et pourquoi les gens adoptent des critères différents dans différents cas, dépendant de la façon dont la situation est présentée ou que l'agent soit un ami ou un étranger, etc. » (Knobe et Doris, 2010, p. 322). [18]

J'aimerais donc proposer qu'une des variables qui affectent le jugement dans les cas d'attribution de responsabilité pour des comportements causés par des attitudes implicites est le fait d'être actuellement ou d'avoir été (probablement sur une base régulière) l'objet de préjugés ou de discrimination. Et en fait, je propose que l'on teste plus avant cette idée en reprenant les expériences de Cameron et ses collègues, en mettant les sujets dans une condition où ceux-ci sont l'objet de la discrimination ou l'ont historiquement été. Ici donc, et dépendant des résultats de l'expérience, on pourrait avoir affaire à une révision que l'on pourrait qualifier de « faible » puisqu'elle s'adresse à la compréhension que nous avons de l'application du concept populaire dans les cas qui nous intéressent. Contrairement à ce que l'on aurait pu penser, nous sommes « variantistes ». Nos jugements sont instables, ils sont affectés par une variable contextuelle (soit le fait d'être ou d'avoir été l'objet de discrimination) et on pourrait peut-être découvrir que la base de nos jugements est différente selon qu'ils portent sur une victime ou sur quelqu'un qui discrimine.[19]

La littérature que nous venons de survoler semble montrer que les victimes de discrimination voient les choses autrement que ceux qui la commettent. On peut interpréter le jugement des premiers de deux façons, une seule de celle-ci supportant le variantisme. La première consisterait à soutenir que les victimes jugent que ceux qui discriminent sont des hypocrites, qu'ils disent une chose, mais en pensent et font une autre (ils sont de mauvaise foi, en quelque sorte). La seconde façon consisterait à soutenir que, pour ceux qui subissent la discrimination, la conscience de ceux qui discriminent n'est pas essentielle au jugement de responsabilité. Le texte de Richeson et Shelton, auquel Hardin et Banaji faisaient allusion, pourrait nous faire pencher pour la première interprétation (selon laquelle le premier groupe ressent les choses comme une forme d'hypocrisie). Mais la seconde interprétation est toujours possible, selon laquelle la base sur laquelle repose les attributions de responsabilité est tout simplement différente d'un groupe à l'autre dans ces situations (la conscience n'est pas requise). Dans ce cas, nous pourrions être justifiés d'attribuer à ceux qui souffrent de discrimination, lorsqu'ils jugent ceux qui discriminent, une théorie de la responsabilité différente de celle qui met la conscience au centre de la responsabilité[20]. De telles théories ont été proposées dans la littérature et j'aimerais en donner un aperçu rapide.

Plusieurs auteurs (Arpaly, 2003 ; Sher, 2006, 2009 ; Smith, 2005, 2008a) ont soutenu que, sur une base quotidienne, nous tenons pour acquis que les gens peuvent être responsables de plus que ce dont ils sont conscients ou de ce qu'ils ont l'intention déclarée de faire : je peux vous tenir responsable d'avoir oublié mon anniversaire, de ne pas être sensible à ce que je vous dis ou bien d'avoir oublié le chien dans la voiture en plein soleil, sans qu'en aucun temps vous ne soyez conscient de le faire ou que vous n'ayez eu l'intention de le faire. Si tel est le cas, les théories comme celles de Levy négligent un élément important de notre concept de responsabilité : le fait que la présence de la conscience ou de l'intention consciente ne pèse pas toujours dans l'attribution de responsabilité.

George Sher pense, de la même façon que proposait Pronin plus haut, qu'il y a une asymétrie dans l'attribution de responsabilité et qu'il n'est pas du tout évident que l'attribution à la première personne se fasse sur la même base que l'attribution à la troisième personne. Comme l'écrit Sher :

> De notre point de vue rétrospectif et externe, [l]es croyances conscientes [de l'agent] n'ont pas de priorité particulière sur sa constitution physique ou sur ses attitudes et traits inconscients. Donc, pour autant que blâmer les gens et les tenir responsables sont des réactions que nous avons à leur égard d'une perspective qui ne coïncide pas avec la leur, la façon dont nous devons voir leurs situations de choix lorsqu'ils délibèrent ne nous procure aucune raison évidente de ne pas baser notre blâme ou notre attribution de responsabilité sur des faits qui les concernent et sur la situation dans laquelle ils ont fait des choix dont ils n'étaient même pas conscients (2009, 61).

Sur quoi donc basons-nous notre jugement alors ? Sher propose une théorie néo-humienne de la responsabilité qui capturerait l'essence de nos jugements (ceux du sens commun) dans les cas de responsabilité pour des actions dont nous ne sommes pas conscients. Elle est néo-humienne parce que, comme Hume avant lui (1739/1995), Sher propose que le lien entre l'agent et l'action reprochée passe par le caractère de l'agent[21]. Ce qui est véritablement blâmé, ce n'est pas tant l'action que le caractère de l'agent dont elle découle[22].

Sher, comme Pronin, soutient une position générale selon laquelle il existe une asymétrie entre les jugements à la première personne et les jugements à la troisième personne. Notre proposition est un peu plus circonscrite. Elle suggère qu'une variable qui affecte le jugement de responsabilité est le fait d'être objet de discrimination. Dans ces conditions, les victimes ne tiennent pas compte des intentions avouées (ou leur donnent moins de poids) de ceux qui discriminent, mais tiennent plutôt compte de leurs actions. J'ai suggéré qu'il est possible d'avoir deux interprétations de la position des victimes : ou bien (1) ils considèrent ceux qui discriminent comme des hypocrites ou comme étant de mauvaise foi ; ou bien (2) ils ne présupposent pas que ceux qui discriminent sont nécessairement conscients de ce qu'ils font et jugent plutôt que leurs actions sont le reflet de leur « caractère ». Selon la première interprétation, les juges des actes de discrimination établiraient la responsabilité sur la base de la présence d'intentions présumées. Selon la seconde interprétation, ils n'auraient que faire de ces intentions, et jugeraient plutôt le caractère des agents sur la base d'actions qui sont particulièrement révélatrices de celui-ci (peut-être parce qu'elles sont interprétées comme une preuve de la mauvaise volonté ou encore de l'indifférence de la part de celui qui les commet)[23]. Cette dernière interprétation est une interprétation plausible des résultats (encore fragmentaires, il convient de le noter) concernant les afro-américains dans l'expérience de Cameron et al. (2010). S'il s'avérait que cette interprétation était la bonne, cela militerait en faveur d'une forme de variantisme du type de celui que proposaient Knobe et Doris (en effet, la première interprétation n'exige pas que nous utilisions des critères différents). Si le variantisme n'est pas compatible avec la façon dont le sens

commun et la plupart des philosophes se représentent la façon dont nous procédons à nos attributions de responsabilité, alors l'acceptation de cette thèse constitue une révision (faible) de la forme que prend la théorie de la responsabilité du sens commun (et potentiellement des théories philosophiques qui s'en inspirent).

4.2 LES THÉORIES DU VRAI MOI

Dans la section précédente, j'ai défendu l'idée que les critères qu'appliquaient les gens pour déterminer la responsabilité étaient « variables » et qu'ils pouvaient pour cette raison supporter la construction de théories radicalement différentes de la responsabilité. Je proposais que dans certains contextes, les théories basées sur les intentions conscientes ne permettaient pas de rendre compte des jugements des gens. Reste qu'il semble que dans certains cas, ces théories font parfois le travail, c'est-à-dire qu'elles capturent une partie des intuitions des gens. Je m'intéresse donc dans cette section au potentiel révisionniste des études sur les attitudes implicites pour un certain type de théorie qui donne une place privilégiée à la conscience.

Dans son « Restoring Control » (2008), Neil Levy, commentant des cas du type de ceux qui nous intéressent, écrit que la responsabilité requiert ce qu'il nomme « l'attributabilité profonde » (*deep attributatibility*). Les actions sont « profondément attribuables » seulement « lorsque la conscience joue un rôle substantiel dans leur production » (217). Et, généralement, cette contribution se fait à travers la délibération, si bien que « la délibération consciente reflète proprement la personne entière, y compris ses valeurs consciemment endossées » (*idem*, 220). Après Susan Wolf, on nomme les théories de ce genre théories du « vrai moi » (parfois également, on parle de théories « identificationnistes »). Selon ce genre de théories, défendues entre autres par Dworkin (1976), Frankfurt (1971) et Watson (1975), un agent n'a de liberté de volonté (et ne peut être tenu pour responsable) que s'il peut s'identifier avec un élément psychologique spécial (que ce soit un désir, une valeur, une volonté ou une intention[24]) qu'il considère comme son vrai moi (par opposition à d'autres désirs, valeurs, volontés ou intentions qu'il considère comme « étrangers à lui-même », faisant partie de son « faux-moi » en quelque sorte). Comme l'écrit Nahmias qui résume assez bien cette position :

> [l]'agentativité autonome [qui est nécessaire, selon certains, pour la responsabilité] requiert l'habileté de former et d'agir selon des principes. La formation de ces principes devrait se produire à travers une délibération consciente exempte de l'influence de motivations inconscientes que l'agent rejetterait s'il en connaissait l'existence (2006, p. 170).

Une objection classique à cette position (entre autres, Arpaly, 2003 ; Arpaly et Schroeder, 1999 ; Thalberg, 1978) consiste à se demander pourquoi on devrait donner un tel privilège à la délibération consciente dans l'établissement de ce qui constitue mon « vrai moi » ou, comme l'écrit Thalberg à propos des théories hiérarchiques de Frankfurt et Dworkin (mais l'objection peut également être appliquée avec quelques modifications aux théories non hiérachiques du type de celle de Watson) :

> Pourquoi tenir pour acquis que l'attitude de second niveau de l'agent doive toujours être plus fondamentalement la sienne, plus représentative de ce qu'il désire fondamentalement, que celle qu'il adopte au premier niveau ? Peut-être que son attitude de niveau supérieur n'est qu'un vague état d'âme qui le ronge[25] (1978, 234).

Il est donc possible de remettre en question mon sérieux quant à ce qui prétendument m'importe (ou ce à quoi j'accorde de la valeur) et, par conséquent, ce que je considère véritablement mien et ce que je rejette comme n'étant pas mien. Dans le contexte de notre discussion, on peut entre autres se demander si la dissociation que révèlent les tests sur les attitudes explicites et les attitudes implicites ne relève pas d'une confusion des sujets par rapport à ce qu'on leur demande. Peut-être, lorsqu'on leur demande ce qu'est leur attitude explicite, ne réfèrent-ils qu'à ce qu'ils croient désirable, plutôt que ce à quoi ils adhèrent vraiment, c'est-à-dire leurs véritables motivations. Il faudrait, semble-t-il, pouvoir répondre aux questions suivantes : Comment savoir si une attitude que je rapporte est vraiment mienne ? Comment savoir qu'une motivation est vraiment la mienne, qu'elle reflète un jugement que j'endosse véritablement ? J'aimerais présenter dans ce qui suit deux ensembles de travaux qui montrent que les méthodes développées dans le domaine des attitudes implicites pourraient aider à répondre à cette question.

La première étude s'est intéressée à la possibilité que les dissociations entre attitudes implicites et attitudes explicites relèvent de préoccupations concernant l'image de soi vis-à-vis d'autrui. L'idée est qu'il est possible que la source de la variation entre ce que révèlent nos attitudes explicites et ce que révèlent nos attitudes implicites provienne du fait que, dans les tests d'attitudes explicites, nous pouvons contrôler nos réponses et vouloir qu'elles peignent de nous une image qui est conforme aux attentes sociales. Ainsi, ces études ne permettraient pas de découvrir nos véritables motivations, ou nos véritables attitudes, mais plutôt celles que nous considérons comme étant positives dans notre société. L'étude en question est celle de Nier (2005), qui vise à contourner ce problème. Ce dernier utilise un stratagème déjà utilisé par Sigall et Page (1971) dans une de leurs études. L'idée est la suivante : on dit à des sujets que l'expérimentateur a en sa possession une machine ou un moyen qui lui permet de détecter quelles sont les véritables attitudes des sujets (dans le cas de Nier, il dit à certains de ces sujets que son test pour détecter les attitudes implicites est un révélateur des véritables attitudes explicites de ceux-ci). Cette procédure, nommée la « pseudo-ligne directe » (*bogus pipeline*), aurait l'avantage de minimiser la force de la désirabilité sociale sur ce que disent les sujets (leur mensonges seront détectés, il vaut mieux dire la vérité que de vouloir bien paraître et passer pour un hypocrite), et c'est ce que confirment les résultats de l'étude de Nier :

> Les résultats indiquent que, lorsque les participants croyaient que leurs « véritables attitudes » étaient évaluées précisément, il y avait une relation significative entre les mesures implicites des attitudes raciales (le test des attitudes implicites) et la mesure explicite des attitudes raciales (l'échelle du racisme moderne). […]

> Ainsi les résultats suggèrent que lorsque la motivation pour rapporter les attitudes explicites qui sont consistantes avec les attitudes implicites augmente, la relation implicite/explicite devient plus forte en raison du changement dans les attitudes explicites rapportées par le sujet […] » (2005, 48).

Plus loin dans son texte, Nier prétend que ces résultats vont à l'encontre de l'idée que la dissociation entre les attitudes provient du fait que les attitudes implicites et les attitudes explicites sont des types de représentations différents qui sont dus à des processus cognitifs différents. Du point de vue de leur accès à la conscience, nos attitudes explicites et nos attitudes implicites pourraient bien être, jusqu'à un certain point, identiques[26].

Ces travaux semblent donc proposer que les attitudes implicites des sujets ne leur sont pas étrangères, que ceux-ci savent qu'ils les possèdent. Étant conscients de ces attitudes, ils devraient donc faire en sorte de tenter de s'en débarrasser ou, au moins, de les neutraliser (le contenu de la note 26 semble indiquer qu'ils ignorent comment les empêcher). Mais on pourrait aussi penser qu'ils seront tentés de les corriger uniquement s'ils sont véritablement motivés pour le faire. Comment dès lors savoir qu'ils possèdent véritablement cette motivation à ne pas endosser leurs attitudes (implicites ou explicites) ?

Ici encore, la recherche sur les attitudes implicites peut nous venir en aide et nous fournir un précieux outil pour nous permettre de trancher la question. Pour comprendre le fonctionnement de cet outil, il importe de comprendre une distinction, qui est établie en psychologie, entre les motivations internes et les motivations externes. Amodio (2008) décrit cette distinction et ses effets ainsi :

> Notre recherche antérieure a récemment montré que les biais raciaux implicites (qui ont déjà été considérés comme inévitables et immuables) des participants varient en fonction de leurs motivations à répondre sans préjugé. C'est-à-dire que les Américains de race blanche rapportent qu'ils répondent sans préjugé aux personnes de race noire pour deux raisons : ou bien pour rencontrer leurs standards personnels ou internes [raisons internes], ou bien pour éviter les réactions négatives des autres qui désapprouvent les préjugés [raisons externes] […]. Devine et al. […] ont montré que les gens qui étaient motivés uniquement par des raisons internes […] présentaient de façon consistante des niveaux de biais implicites plus bas dans les mesures de temps de réaction que les participants qui rapportaient d'autres profils motivationnels [c.-à-d., qui n'étaient motivés que par des raisons internes ou par un mixte de raisons internes et externes] » (12).

Dans ce contexte, il devient important de pouvoir mesurer ces motivations internes. Comment savoir si nous sommes véritablement motivés de façon interne ? La question se pose, puisque comme on l'a vu plus haut, il est toujours possible, à cause des pressions sociales, de prétendre être motivés par des raisons internes alors que nous sommes motivés par des raisons externes. Ce sont donc motivations internes qu'il faudrait avoir accès pour mesurer notre motivation.

Comment avoir accès aux motivations internes en étant sûr qu'elles ne sont pas une façade ? Une façon consiste à mesurer les motivations implicites en supposant que celles-ci reflètent les véritables motivations du sujet (c'est-à-dire qu'elles ne sont pas infectées par le souci du bien paraître ou par). Mais comment mesurer ces motivations implicites ? Dans « Implicit Motivation to Control Prejudice » (2008), Glasser et Knowles proposent un moyen de les mesurer. Selon eux, la motivation implicite à contrôler les préjugés (MICP) est composée de deux éléments : d'une attitude négative vis-à-vis des préjugés (ANP) et d'une croyance que j'entretiens des préjugés (CEP). Il est possible de mesurer chacune des composantes à l'aide des tests habituels de mesure des attitudes implicites. Pour la composante ANP, le test consiste à mesurer la force de l'association entre des mots comme « tolérance » ou bien « préjugés », avec des mots comme « bon » ou « mauvais » (ou des équivalents). Pour la composante CEP, le test consiste à mesurer la force de l'association entre des mots comme « tolérant » et « biaisé », et « moi » ou « étranger » (ou des équivalents). Ce que révèlent ces tests est surprenant et intéressant.

> Ceux qui ont des scores élevés au CEP et élevés à l'ANP sont les seuls à montrer une relation non positive (en fait, légèrement négative) entre les stéréotypes associés aux armes et le biais du tireur. Ceux qui ont des scores bas à l'ANP et élevés au CEP montre la relation positive la plus forte entre les stéréotypes liés aux armes et le biais du tireur, reflétant peut-être la motivation implicite d'*utiliser* les stéréotypes (c'est-à-dire, « il est correct d'avoir des préjugés » [bas ANP], « j'ai des préjugés » [CEP élevé], « je vais donc les utiliser ! ») (2008, p. 19).

Autrement dit, ce que montrent Glasser et Knowles, c'est que ceux qui ont une MICP élevée ne montrent pas les effets terribles et dévastateurs qui sont caractéristiques de ceux qui possèdent les attitudes implicites raciales dont nous parlions plus haut. Dans le cas qui nous occupe, la présence du stéréotype selon lequel les noirs sont plus à risque d'avoir des armes que les blancs ne les conduit pas à tirer plus rapidement sur les noirs que sur les blancs (ou à distinguer plus rapidement une arme d'un outil) et à faire plus d'erreurs comme celle de prendre un outil pour une arme.

En résumé, les études que nous venons de considérer nous montrent deux choses : d'abord, que nos attitudes implicites ne sont peut-être pas sous le radar de la conscience, qu'elles ne sont peut-être pas des corps étrangers dont nous ignorerons l'existence. Ce fait ne demande pas de révision à notre concept de responsabilité, il demande plutôt une révision à notre façon de comprendre les expériences produites dans le cadre des théories duales des attitudes. Si ces attitudes sont conscientes, le fait que nous ne les rejetons pas de tout cœur, pourrait bien, selon les critères des théoriciens du vrai moi, mais aussi selon le sens commun, nous rendre potentiellement responsables de celles-ci. C'est peut-être ainsi que l'on peut interpréter les jugements des sujets de l'expérience de Cameron et ses collègues concernant la vignette où l'on présente un Donald aux prises avec une attitude automatique, mais consciente. Donald est conscient de

son attitude, mais malgré le fait qu'il dise ne pas vouloir l'endosser, il ne fait rien pour l'empêcher, prouvant ainsi qu'il ne la rejette pas de tout cœur. Mais comment savoir si nous rejetons réellement ces attitudes ? Après tout, ce n'est que lorsque nous possédons réellement une motivation à ne pas agir sur ces attitudes que nous pouvons les neutraliser. Ici, et c'est la deuxième leçon que nous offrent les travaux que nous avons considérés, la psychologie nous aura fourni un moyen de trancher en ce qui concerne les motivations qui sont vraiment les « miennes », ce qui posait un problème à la théorie du vrai moi. La révision que je propose est donc la suivante : quand on voudra connaître quelles sont les véritables motivations d'un individu, il ne faudra pas se fier à ce qu'il dit (puisque nous pouvons toujours avoir affaire à une position de façade), mais plutôt sonder ses motivations implicites. Ce ne sont qu'elles qui constituent des indices fiables de ce que nous sommes véritablement ! Cette révision n'est donc pas une révision de concept de responsabilité comme telle, mais plutôt une révision de nos pratiques d'attribution de responsabilité et elle vaut pour ceux qui sont à la recherche du véritable moi !

4.3 LA NOTION DE CONTRÔLE

La conception classique de responsabilité depuis Aristote fait du contrôle une des conditions nécessaires de celle-ci. Pour certains, le contrôle implique l'idée de la conscience, de telle sorte qu'il est impossible de penser que le contrôle puisse être inconscient et donc que la condition posée par Aristote (et dont nous avons parlé plus haut) puisse être satisfaite en l'absence de conscience. Or, si tel est le cas, les perspectives sont assez sombres quant à la responsabilité. Comme l'écrivent Hardin et Banaji :

> La recherche montre qu'il est pratiquement impossible de corriger consciemment les effets des préjugés implicites […]. Pour ce faire, on doit être dans la circonstance improbable où l'on a à la fois : a) une connaissance du fait que le préjugé implicite opère, b) avoir une motivation et la capacité cognitive de le contrôler, et peut-être, le plus improbable de tous, c) avoir une connaissance précise de la magnitude et de la direction de la correction nécessaire » (sous presse, 7 ; voir également Nahmias, 2006).

Cela conforterait bien sûr ceux qui pensent que nous ne sommes pas responsables de nos attitudes implicites (et ceux qui, comme Bargh, tirent de ces observations l'implication terriblement déprimante dont nous parlions plus haut). Mais ce serait peut-être juger un peu trop rapidement ce qu'exige le contrôle de notre comportement.

La littérature sur le contrôle suggère qu'il n'est pas nécessaire que les trois conditions stipulées par Hardin et Banaji soient présentes pour qu'il y ait contrôle. En fait, il y a même peut-être plus de chance que le contrôle fonctionne si certaines de ces conditions ne sont pas satisfaites. En effet, comme le notent Pearson et ses collègues :

> Alors que les efforts conscients pour éviter les stéréotypes peuvent souvent échouer ou même exacerber ceux-ci parce que les individus ne connaissant pas

> les processus qui les favorisent et les régulent, des buts implicites passifs pourraient bien réussir en cooptant les processus psychologiques mêmes qui […] sous-tendent [ces stéréotypes] et remplacent les associations stéréotypiques lorsque nous percevons les membres d'un autre groupe racial ou ethnique, ou interagissons avec eux (2009, 16).

J'aimerais présenter deux façons de contrôler les attitudes implicites qui n'exigent pas qu'on soit conscient qu'elles agissent, ni qu'on soit motivé de les appliquer[27] ou de connaître la magnitude du contrôle que l'on devrait exercer. En gros, ces méthodes ne demandent pas qu'on reconnaisse que les stéréotypes agissent, ni une connaissance des mécanismes qui permettent de les contrecarrer.

On distingue dans la littérature deux formes de contrôle : le contrôle « descendant » (*top-down*) et le contrôle « ascendant » (*bottom-up*). On distingue ainsi les formes de contrôle en fonction de leurs réquisits. Le contrôle descendant est le contrôle que manifeste le sujet qui décide consciemment, devant une instance de discrimination, de ne pas laisser ses attitudes contrôler son comportement. Le contrôle ascendant est un contrôle d'une forme différente, qui ne dépend pas de la reconnaissance de l'activation des stéréotypes, ni d'une intention d'inhiber son comportement au moment où le stéréotype est reconnu. Dans ce qui suit, je vais présenter deux formes de ce contrôle ascendant : « l'implémentation d'intention » et « l'amorçage (*priming*) de buts égalitaires ».

Tout d'abord, l'implémentation d'intention est un procédé proposé par Stewart et Payne (2008) dans lequel on propose aux sujets de répéter des phrases qui mentionnent un but, celui d'avoir une attitude contre-stéréotypique. Ce but est également lié à un indice qui indique à quel moment le but devrait être activé. Par exemple, ils demandent aux sujets de répéter la phrase « lorsque je vois le visage d'un homme noir, je pense "sécurité" ». Comparant les résultats obtenus en utilisant ce procédé aux résultats obtenus avec des phrases semblables qui ne mettent pas l'accent sur la réduction de l'attitude (par exemple, « lorsque je vois le visage d'un homme noir, je pense précision [ou rapidité]), les auteurs ont montré que les sujets dans la première condition ne montraient plus les effets caractéristiques d'un « biais lié aux armes », mais un effet contraire, c'est-à-dire qu'ils prenaient plus de temps pour reconnaître une arme dans les mains d'un Afro-Américain[28].

La seconde technique, l'amorçage de buts égalitaires, est proposée par Montieth et ses collègues dans « Schooling the Cognitive Monster » (2009). Dans cet article, les auteurs tentent d'examiner si des « buts égalitaristes qui ne sont pas entretenus de façon chronique peuvent mener au contrôle de l'activation des stéréotypes en déclenchant des opérations ayant pour but de s'en protéger, y compris de se protéger de leur inhibition » (105). S'inspirant de l'idée développée dans la littérature sur la sélection des buts, selon laquelle la contemplation d'un échec dans un domaine active le but relatif à ce domaine, Montieth et ses collègues demanderont aux sujets de réfléchir à une expérience d'échec dans deux domaines différents : d'abord, dans celui de l'action égalitariste, et dans

celui du respect des traditions. Dans chacun des cas, les sujets sont invités à se remémorer une expérience où ils ont échoué ou bien à traiter quelqu'un de façon égale et équitable, ou bien à respecter une tradition. Ce que révèle l'expérience, c'est que dans le cas où on se remémore un échec dans le domaine égalitariste, par opposition à un échec dans un autre domaine ou à une réussite dans le même domaine, un contrôle des stéréotypes s'effectue, contrôle qui se traduit par le fait qu'après avoir vu le visage d'un homme de race noire, les sujets ne répondent pas plus rapidement aux mots qui sont en accord avec les stéréotypes qu'aux autres, et même ils y répondent plus lentement.

À la lumière des travaux précédents, je propose donc la révision suivante : des processus inconscients peuvent être en charge de formes significatives de contrôle. C'est d'ailleurs ce que proposaient récemment Suhler et Churchland (2009) qui écrivaient : « des processus inconscients peuvent soutenir une forme robuste de contrôle et, par extension, […] la conscience n'est pas une condition nécessaire du contrôle » (341). Par conséquent, si le simple fait de voir une personne agir de façon égalitaire est suffisant pour amorcer des buts égalitaires qui contrôleront le comportement, et ce, sans que le sujet soit conscient d'agir selon ces buts ni de l'influence de ce qu'il voit sur son action, il est possible de parler de *contrôle* inconscient[29]. On pourrait dire que ces mécanismes sont donc sensibles à certaines raisons morales ou à certains aspects moraux de l'environnement. Une fois activés, ils font en sorte de guider le comportement, comme le ferait un mécanisme dont le but aurait été sélectionné par le biais d'un raisonnement pratique conscient.

On peut toutefois se demander si ce genre de contrôle inconscient est suffisant pour établir la responsabilité. Il n'est pas certain que le commun des mortels accepte de considérer comme responsables ceux qui agissent à partir de ce genre de buts activés par l'environnement, mais pour certains théoriciens de la responsabilité ce genre de contrôle inconscient ne pose a priori pas de problème. Par exemple, Fisher et Ravizza écrivent :

> [S]elon notre théorie, un agent est moralement responsable d'une action pour autant qu'elle provienne de son propre mécanisme répondant de façon modérée aux raisons [le mécanisme est « modéré » parce qu'il ne répond pas toujours aux raisons, mais il les reconnaît et y répond assez souvent]. La réponse modérée aux raisons est définie par rapport à la sensitivité du type de mécanisme qui donne lieu à l'action. Nulle part il est requis que ce type de mécanisme soit « le raisonnement pratique ». Et nous soutenons que *notre théorie s'applique naturellement et sans heurt aux mécanismes non réflexifs de divers types* » (1998, 86 ; nous soulignons).

Si nous acceptons ce que nous avons présenté à la section précédente, c'est-à-dire l'idée que nous puissions avoir des motivations implicites et que celles-ci sont nos véritables motivations, et si nous acceptons les résultats des études de Payne et de Monteith, il semble qu'il soit possible d'activer ces motivations de façon à ce qu'elles contrôlent notre comportement sans nécessairement que nous

ayons à « nous en mêler » consciemment. Les difficultés qu'évoquaient Hardin et Banaji ne semblent donc pas nous mener à accepter l'implication terriblement déprimante de Bargh. Ainsi, si l'implication terriblement déprimante était la conséquence d'une conception de la responsabilité où le contrôle ne peut être que volontaire et conscient, alors ce que nous venons de présenter semble forcer une révision quant à une interprétation de la condition du contrôle : celui-ci n'a pas à être conscient. Notre révision est donc encore une fois faible, puisqu'elle ne remet pas en cause la condition classique du contrôle, mais qu'elle remet plutôt en cause une de ses interprétations (que je pense cependant être l'interprétation classique. En effet, je doute qu'Aristote incluait le contrôle inconscient dans les formes du contrôle auxquelles il pensait).

5. CONCLUSION

J'ai soutenu que certaines révisions de notre façon de comprendre la responsabilité et nos pratiques d'attribution de la responsabilité pourraient être justifiées par les travaux récents sur les attitudes implicites. M'inspirant des distinctions introduites par Vargas, j'ai distingué une forme de révisionnisme global de formes plus parcellaires ou plus locales. C'est à une forme de révisionnisme parcellaire, à la pièce, que j'ai invité le lecteur. Je n'ai donc pas soutenu que notre concept de responsabilité devait être abandonné ou révisé *in toto* à la lumière des travaux en psychologie sociale, mais plutôt que certaines révisions modestes (pour l'instant) pouvaient être inspirées par ces travaux. Au terme de ce texte, je ne peux pas être sûr d'avoir réussi à convaincre le lecteur d'adopter les révisions que je propose (ou même d'adopter une forme de révisionnisme tout court), mais je pense avoir au moins réussi à montrer deux choses : d'une part, que l'existence des attitudes implicites pose un sérieux problème aux théories de la responsabilité ; et d'autre part, que certains travaux sur ces attitudes nous permettent d'affiner notre réflexion sur la question de notre responsabilité vis-à-vis leurs effets.

Finalement, et même si ce n'était pas l'objectif central de mon texte, je crois également avoir montré que la menace constituée par le monstre cognitif (et les implications terriblement déprimantes que Bargh tirait de son existence) a été surestimée (même si elle ne doit pas pour autant être sous-estimée). Nous ne sommes pas des marionnettes dont le comportement est déterminé par les caractéristiques de notre environnement[30].

BIBLIOGRAPHIE

Amodio, D. M. 2008. « The Social Neuroscience of Intergroup Relations ». *European Review of Social Psychology, 19*, 1-54.

Amodio, D. M. et P.G. Devine. 2006. « Stereotyping and Evaluation in Implicit Race Bias: Evidence for Independent Constructs and Unique Effects on Behavior ». *Journal of Personality and Social Psychology, 91*, 652-661.

Arpaly, N. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. New York: Oxford University Press.

Arpaly, N, et T.Schroeder. 1999. « Praise, Blame and the Whole Self ». *Philosophical Studies*, 93:2, 161-188.

Banaji, M. R., B. A. Nosek et A. G. Greenwald. 2004. « No Place for Nostalgia in Science: A Response to Arkes and Tetlock ». *Psychological Inquiry*, vol. 15, no. 4, 289-289.

Bargh, J. A. (1999). « The Cognitive Monster: The Case Against Controllability of Automatic Stereotype Effects ». In S. Chaiken & Y. Trope (dir. publ.), *Dual Process Theories in Social Psychology*. New York: Guilford, 361-382.

Bargh, J.A. et T. L. Chartrand.1999. « The Unbearable Automaticity of Being ». *American Psychologist*, 54, 462-479.

Bargh, J.A. et E. Morsella. 2008. « The Unconscious Mind ». *Perspectives in Psychological Science*, 3, 73-79

Bruneau, E. G., N. Dufour, et R. Saxe. 2012. « Love, Hate and Indifference: Behavioral and Neural Responses in Arabs, Israelis and South Americans to Each Others' Pain and Suffering. Philosophical Transactions of the Royal Society: Biology, 367, 717-730.

Cameron, C. D., K. Payne et J. Knobe. 2010. « Do Theories of Implicit Race Bias Change Moral Judgments? ». *Social Justice Research*, 23, 272-289

Correll, J., B. Park, C. M. Judd, and B. Wittenbrink. 2002. « The Police Officer's Dilemma: Using Race to Disambiguate Potentially Threatening Individuals ». *Journal of Personality and Social Psychology* 83:1314-29.

Dworkin, R. 1976. « Autonomy and Behavior Control ». *Hasting Center Reports*, 6 :1, 23-28.

Dunham, Y. A. S. Baron, M. R. Bnanji 2008. « The Development of Intergroup Cognition ». *Trends in Cognitive Sciences*, 12, 7, 248-253.

Evans, J. (2003). « In Two Minds: Dual-Process Accounts of Reasoning ». *Trends in Cognitive Sciences*, 7: 10, 454-459.

Faucher, L. et E. Machery. 2009. « Racism: Against Jorge Garcia's Moral and Psychological Monism ». *Philosophy of the Social Sciences*, 39: 1, 41-62.

Ferguson, M. J. et J. Bargh, 2004. « How Social Perception Can Automatically Influence Behavior ». *Trends in Cognitive Sciences*, 8: 1, 33-39.

Fisher, J. M. et M. Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. 1971. « Freedom of the Will and the Concept of a Person ». *Journal of Philosophy*, 68, 5-20.

Frankish, K. 2010. « Dual-Process and Dual-System Theories of Reasoning ». *Philosophy Compass*, 5:10, 914-926.

Frankish, K. et J. Evans. 2009. « The Duality of Mind: An Historical Perspectives ». In Evans, J. et K. Frankish (dir.publ.), *In Two Minds: Dual Processes and Beyond*, Oxford: Oxford University Press, 1-29.

Gawronski, B. et L. A. Creighton. Sous presse. « Dual-Process Theories ». In D. E. Carlston (dir. publ.), *The Oxford Handbook of Social Cognition*, N.Y.: Oxford University Press.

Glasser, et E.D. Knowles (2008). « Implicit Motivation to Control Prejudice ». *Journal of Experimental Social Psychology*, 44, 164-172.

Greenwald, A. G. et M. R. Banaji. 1995. « Implicit Social Cognition: Attitudes, Self-Esteem and Stereotypes ». *Psychological Review*, 102, 4-27.

Haidt, J. 2001. « The Emotional Dog and its Rational Tail ». *Psychological Review*, 108, 814-834.

Hardin, C.D. et M. Banaji. Sous presse. « The Nature of Implicit Prejudice: Implications for Personal and Public Policy ». In E. Shapir (dir.publ.), *The Behavioral Foundations* of Policy.

Hobbes, T. 1654/1977. *De la liberté et de la nécessité*. Paris : Vrin.

Hume, D. 1739/1995. *Traité de la nature humaine*. Paris : Flammarion.

Kahneman, D. 2002. « Maps of Bounded Rationality: A Perspective on Intuitive Judment and Choice ». *Nobel Prize Lecture* (http://www.nobelprize.org/nobel_prizes/ economics/laureates/2002/kahnemann-lecture.pdf)

Kelly, D., L. Faucher et E. Machery. 2010. « Getting Rid of Racism: Assessing Three Proposals in Light of Psychological Evidence ». *Journal of Social Philosophy*, 41:3, 293-322.

Knobe, J. et J. Doris. 2010. « Responsibility ». In J. Doris and the Moral Psychology Research Group (dir. publ.), *The Moral Psychology Handbook*, New York: Oxford University Press, 321-354.

Levy, N. 2008. « Restoring Control: Comments on George Sher ». *Philosophia*, 36: 2, 213-221.

Libet, B., C. A. Gleason, E. W. Wright et D.K. Pearl, D. K. 1983. « Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act ». *Brain*, 106: 623-642.

Machery, E., L. Faucher et D. Kelly. 2010. « On the Alleged Inadequacy of Psychological Explanations of Racism ». *The Monist*, 93:2, 228-254.

Mendoza , S. A., P. M. Gollwitzer et D.M. Amodio. 2010. « Reducing the Expression of Impli-

cit Stereotypes: Reflexive Control Through Implementation Intentions ». *Personality and Social Psychology Bulletin, 36*, 512-523.

Monteith, M.J., J.E. Lybarger, A. Woodcock. 2009. « Schooling the Cognitive Monster: The Role of Motivation in the Regulation and Control of Prejudice ». *Social and Personality Compass*, 3, 211-226.

Nahmias, E. 2006. « Autonomous Agency and Social Psychology ». In M. Marrafa, M. Caro et F. Ferretti (dir. publ.), *Cartographies of the Mind, Philosophy and Psychology in Intersection*, New York: Springer, 169-186.

Nelkin, D. 2005. "Freedom, Responsibility, and the Challenge of Situationism," in Free Will and Moral Responsibility, Midwest Studies in Philosophy 29. Cambridge: Blackwell. (2005): 181- 206.

Nichols, S. 2007. « After Incompatibilism: A Naturalistic Defense of Reactive Attitudes ». *Philosophical Perspectives*, 21, 405-428.

Nier, J. 2005. « How Dissociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach ». *Group Processes Intergroup Relations*, 8, 1, 39-52.

Nozek, B., C. Hawkins et R. Frazier. 2011. « Implicit Social Cognition: From Measures to Mechanisms ». *Trends in Cognitive Sciences*, 15 (4), 152-159.

Olson, K.R. & Dunham, Y.D. 2010. « The Development of Implicit Social Cognition ». In B. Gawronski et K. Payne (dir.publ.). *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. New York: Guilford, 241-254.

Payne, B. K. 2001. « Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon ». *Journal of Personality and Social Psychology*, 81, 181-192.

Payne, B. K. 2006. « Weapon Bias: Split Second Decisions and Unintended Stereotyping ». *Current Directions in Psychological Science*, 15, 287-291.

Pearson, A., J' Dovidio et S. Gaertner. 2009. « The Nature of Contemporary Prejudice: Insights from Aversive Racism ». *Social and Personality Psychology Compass*, 3, 1-25.

Pizarro, D. A. et D. Tannenbum. 2011. « Bringing Character Back: How the Motivation to Evaluate Character Influences Judgments of Moral Blame ». In Mikulincer, M. et P. Shaver (dir. publ.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. Washington, D.C.: APA Press, 91-108.

Pronin, E. 2007. « Perception and Misperception of Bias in Human Judgment ». *Trends in Cognitive Sciences*, 11, 37-43.

Pereboom, D. 2001. *Living without Free Will*. Cambridge University Press.

Ross, L., & Nisbett, R. E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.

Sher 2009 *Who Knew? Responsibility Without Awareness*, Oxford University Press, forthcoming 2009.

Sher, G. 2006. « Out of Control ». *Ethics*, 116:2, 285-301.

Sigall, H. et R.A. Page. 1971. « Current Stereotypes: A Little Fading A Little Faking ». *Journal of Personality and Social Psychology*, 16, 252-258.

Smilansky, S. 2002. « Free Will, Fundamental Dualism and the Centrality of Illusion ». In R. Kane (dir.publ.), *The Oxford Handbook of Philosophy*, Oxford University Press, p. 489-505.

Smith, A. 2005. « Responsibility for Attitudes: Activity and Passivity in Mental Life ». *Ethics* 115 (2005), pp. 236-271

Smith, A. 2008a. « Control, Responsibility, and Moral Assessment, » *Philosophical Studies*, 138,367-392.

Smith, A. 2008b. « Character, Blameworthiness, and Blame: Comments on George Sher's, *In Praise of Blame* ». *Philosophical Studies* 137, 31-39.

Smith, E. R. et E. C. Collins. 2009. « Dual-Process Models: A Social Psychological Perspective». In J. Evans et K. Frankish (dir. publ.), *In Two Minds: Dual Processes and Beyond*, New York: Oxford University Press, 197-216.

Spinoza, 1677/1994. *L'Éthique*. Paris : Gallimard.

Stewart, B.D., & Payne, B.K. 2008. « Bringing Automatic Stereotyping Under Control: Implementation Intentions as Efficient Means of Thought Control ». *Personality and Social Psychology Bulletin*, 34, 1332-1345

Suhler, C.L. and Churchland, P.S. 2009. « Control: Conscious and Otherwise ». *Trends in Cognitive Sciences*, 13(8), 341-347

Sunstein, C. R. 2005. « Moral Heuristics ». *Behavioral and Brain Sciences* 28 (4):531-542.

Strawson, P. 1962. « Freedom and Resentment ». *Proceedings of the British Academy*, 48, 1-25.

Thalberg, I. 1978. « Hierachical Analyses of Unfree Action ». *Canadian Journal of Philosophy*, vol. 8 : 2, 211-226.

Uhlmann, E et B. Nosek. Sous presse. « My Culture Made Me Do It : Lay Theories of Responsibility for Automatic Prejudice ». *Social Psychology*.

Unkelbach, Christian ; Forgas, Joseph ; Denson, Thomas (2008). « The Turban Effect : The Influence of Muslim Headgear and Induced Affect on Agressive Responses in the Shooter Bias Paragidm ». *Journal of Experimental Social Psychology* 44 (5): 1409–1413.

Vargas, M. 2004. « Responsibility and the Aims of Theory: Strawson and Revisionism ». *Pacific Philosophical Quarterly* 85:2, 218-241.

Vargas, M. 2005. « The Revisionist's Guide to Responsibility ». *Philosophical Studies* 125:3,399-429.

Vargas, M. à paraître. « Situationism and Responsibility: Free Will in Fragments ». In T. Vierkant, J. Kiverstein, et A. Clark (dir.publ.), *Decomposing the Will*.. New York: Oxford University Press.

Watson, G. 1975. « Free Agency ». *Journal of Philosophy*, 72: 8, 205-220.

Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wellman, D. 2007, « Unconscious Racism, Social Cognition Theory, and the Legal Intent Doctrine: The Neuron Fires Next Time ». In H. Vera et J. R. Feagin (dir. publ.), *Handbooks of the Sociology of Racial and Ethnic Relations*, New York: Springer, 39-65.

Wigley, S. 2007. « Automaticity, Consciousness and Moral Responsibility ». *Philosophical Psychology*, 20:2, 209-225.

Wilson, T. D., S. Lindsey et T. Schooler. 2000. « A Model of Dual Attitudes ». *Psychological Review*, vol. 107, no. 1, 101-126.

Wolf, S. 1993. « The Real Self View ». In J. Fischer and M. Ravizza (dir.publ.), *Perspectives on Moral Responsibility*, Ithaca: Cornell University Press, 151-169.

NOTES

1 Dans ce qui suit, je présente une version volontairement simplifiée des théories des processus duaux. Ces théories prennent plusieurs formes (certaines postulent l'existence de systèmes différents regroupant des processus partageants certaines propriétés ; d'autres proposant des rapports différents entre les systèmes) et, pour des raisons d'espace, je ne présenterai que ce que je considère être le noyau commun des thèses de ces théories. Pour une discussion plus complète, voir Frankish (2010).

2 Encore une fois, je présenterai une version simplifiée des théories des attitudes duelles. Je présenterai ce que l'on désigne sous le nom de « modèle des attitudes duelles » qui me semble être le modèle le plus fréquemment utilisé dans la littérature (Wilson et al., 2000). Est-il nécessaire de le préciser, ce modèle n'est pas le seul à vouloir expliquer les phénomènes qui intéressent les psychologues sociaux et ne fait pas consensus (voir Gawronski et Creighton pour une présentation des différents modèles).

3 Dans ce qui suit, j'utiliserai le terme « attitude » pour référer à ce groupe hétérogène, même si certains des psychologues distinguent les états cognitifs (comme les stéréotypes ou biais) des attitudes. Selon cette distinction, on pourrait avoir une attitude positive (dans ce cas, ce à quoi l'on réfère est une disposition affective positive) vis-à-vis des femmes, mais avoir des stéréotypes négatifs à leurs égards (Banaji et al., 2004). Dans le contexte de notre discussion, ma préférence pour le terme « attitude » provient de ce que, comme Amodio et Devine (2006), je pense que les psychologues n'ont souvent pas réalisés (ou ils n'ont pas été capables de préciser lequel des constituants pouvait fournir une explication) que les résultats de leurs tests de mesure sont expliqués parfois par les états affectifs implicites et parfois par les états cognitifs implicites (ou par la combinaison ou la disjonction des deux types), je choisis dans le contexte de l'article de référer à ces états en utilisant un vocable unique.

4 Comme le notent Wilson et ses collègues, les attitudes duelles sont distinctes de l'ambivalence en ce que les sujets qui les possèdent ne font pas l'expérience subjective d'un conflit ou d'une division de l'esprit, mais rapporte l'attitude qui est la plus accessible.

5 Greenwald et Banaji (1995) définissent les attitudes implicites ainsi : elles sont « des traces d'expériences passées non identifiées de façon introspective (ou identifiées de façon incorrecte) qui contribuent aux sentiments favorables ou défavorables vis-à-vis de l'objet de l'attitude » (6). Nous reviendrons sur la question de savoir si les gens sont conscients ou non de la présence de ces attitudes, mais disons pour l'instant que la majorité des psychologues sociaux caractérisent les attitudes concernant certains groupes sociaux, tels les races, comme inconscientes. Pour les attitudes implicites dont les gens sont conscients, on dira souvent que ce qui les distinguent des attitudes explicites est l'ignorance de l'origine de l'attitude. En d'autres mots, les gens qui ont une attitude implicite ne savent pas pourquoi ils évaluent l'objet cible de la façon dont ils le font.

6 Contrairement aux attitudes explicites (qui peuvent être mesurées à l'aide de questionnaires), les attitudes implicites doivent être mesurées en utilisant des méthodes indirectes (comme l'amorçage évaluatif séquentiel [*sequential evaluative priming*], la procédure de l'attribution affective erronée [*affect misattribution procédure*], etc.). Une de ces mesures, la plus répandue selon Nozek et al. (2011), est une version de la tâche de Stroop où l'on montre à des sujets, sur un écran d'ordinateur, des photos de visages d'hommes blancs ou d'hommes noirs suivies de mots à valence positive (par exemple, « naissance ») ou négative (par exemple, « mort »). Dans une des conditions, la tâche des sujets consiste à appuyer sur un bouton à droite si un visage noir est associé à un mot à valence positive et à gauche si un visage blanc est associé à un mot à valence négative, et dans une autre condition, à faire l'inverse, c'est-à-dire à appuyer sur le bouton de droite si le visage noir est associé à un mot à valence négative et à gauche si le visage blanc est associé à un mot à valence positive. Les différentes latences dans les temps de réponse sont une indication de la force des associations entre les concepts raciaux et les évaluations positives ou négatives. Par exemple, les sujets blancs prennent moins de temps à appuyer sur le bouton de droite quand le visage noir est associé

à un mot à valence négative plutôt qu'à un mot à valence positive, et plus de temps à appuyer sur le bouton de gauche quand le visage blanc est associé à un mot à valence négative plutôt qu'à un mot à valence positive.

7 Payne soutient également que ses résultats s'étendent aussi aux situations où les ressources cognitives et attentionnelles des sujets sont limitées ou affaiblies (à cause de la fatigue ou de l'alcool, par exemple).

8 Je présente ici une version selon moi plus écologiquement valide de l'expérience que Payne a fait subir à ses sujets (2001), soit celle de Correll et al., (2002).

9 C'est d'ailleurs l'inférence que fait Bargh : « Comment pourrait-on être tenu responsable, légalement ou autrement, de comportements discriminatoire ou préjudiciable lorsque la science psychologique a montré que de tels effets se produisent de façon non intentionnelle » (1999, 363).

10 Même si nous n'étions que parfois sous l'influence des attitudes implicites, notre capacité à agir de façon responsable apparaîtrait comme beaucoup plus fragile et limitée que ce que nous pouvons croire. Comme le notait Vargas au sujet de la littérature situationniste (mais son verdict vaut également pour le cas de la littérature qui nous intéresse), « la recherche psychologique suggère que ce qui nous apparaît comme une capacité générale de répondre aux raisons est en fait un groupe de capacités spécifiques, écologiquement limitées et indexées aux situations particulières » (sous presse).

11 Une telle préoccupation est évidente chez Strawson, par exemple, lorsqu'il écrit que son projet est de tenter de rendre compte de « ce que l'on veut dire, c'est-à-dire, ce que nous voulons tous dire » par responsabilité (1962, p. 78 ; ma traduction).

12 « Une idée centrale du révisionnisme est qu'une théorie adéquate de la responsabilité va s'éloigner de façon significative de notre compréhension du sens commun de la responsabilité » (Vargas, 2004, 219).

13 « Si une révision du concept de responsabilité morale implique que le concept révisé ne joue pas la même sorte de rôle (comme concept moral) dans notre réseau de normes, alors la révision a échoué à rencontrer le standard d'adéquation normative » (2004, 231).

14 S'il s'avérait que ce que les philosophes qui prétendent capturer notre concept populaire nous proposaient un concept qui ne correspond pas à celui que la plupart des gens entretiennent, alors il faudrait opérer une révision faible du concept (une révision à notre compréhension de ce qu'est le concept du sens commun). Si jamais il s'avérait que, pour une raison ou pour une autre, le concept du sens commun est inacceptable, il faudrait soit l'éliminer, soit procéder à une révision modérée.

15 Wellman (2007, p. 50) fait également référence à l'absence de contrôle pour excuser les agents.

16 Notons cependant nous n'acquérons probablement pas toutes nos attitudes en bas âge. Si bien qu'une question importante pour l'attribution de responsabilité pourrait bien être de savoir lesquelles de nos attitudes sont acquises en bas âge et lesquelles sont acquises plus tard.

17 Une explication possible de cet effet pourrait être que le fait d'être discriminé a pour effet de faire en sorte de rendre évident le fait que nous n'appartenons pas à un certain groupe et donc à voir plus clairement les actions de ce groupe ou à voir plus clairement l'effet de ces actions (par exemple, en posant une limite à la tendance que nous avons à nous mettre dans les souliers des autres pour interpréter les comportements), voir Bruneau et al. 2012 pour une description d'un des mécanismes potentiels derrière cet effet.

18 Ou encore, « une théorie est invariantiste si elle applique le même critère dans tous les cas où les gens font des jugements concernant la responsabilité morale. Ainsi, une théorie invariantiste pourrait dire : (1) « Quel que soit la personne qui juge, quelque que soient les circonstances, faites toujours les jugements de responsabilité morale en vérifiant si l'agent rencontre les critères suivants … [et, ce serait] un rejet de l'invariantisme de dire que : « (2) « Si l'agent est un ami, utilisez le critère suivant … mais si l'agent est un étranger, utiliser tel autre critère qui est légèrement différent … » (idem, 323). Une conséquence du varian-

tisme est que les jugements des gens quant à la responsabilité peuvent donner un support à des théories philosophiques (qui prétendent capturer les intuitions du sens commun) différentes et parfois contradictoires.

19 Comme me l'a fait remarquer un des évaluateurs de cet article, c'est une chose de dire que nos jugements varient selon que l'on appartient à un groupe ou à un autre (énoncé descriptif), c'en est une autre de dire que c'est une bonne chose qu'ils varient ainsi (énoncé normatif). Je n'ai pas d'autres ambitions ici que de tenter de donner une meilleure description de notre pratique.

20 Dans ce cas, il faudrait vérifier s'ils n'ont qu'*une seule* théorie de la responsabilité qui est différente de la théorie qui requiert la conscience ou bien s'ils n'entretiennent pas simultanément *deux* théories, une pour le groupe auquel ils appartiennent et une pour le groupe qui discrimine. Il me semble possible de vérifier empiriquement cette question en étudiant les conditions de l'attribution de responsabilité aux membres de son propre groupe. On pourrait ainsi observer si les membres des groupes discriminés utilisent les mêmes critères lorsqu'ils attribuent la responsabilité aux membres de leur propre groupe que lorsqu'ils attribuent la responsabilité à ceux qui les discriminent (ou à ceux qui discriminent en général).

21 Les psychologues Pizarro et Tannenbum (2011) soutiennent également que nos jugements de responsabilité reposent sur des jugements concernant le caractère des agents. Si la thèse de ces derniers s'avérait être exacte, il faudrait expliquer différemment la variabilité que nous observons entre les discriminés et ceux qui discriminent. L'explication prendrait la même forme que celle que je présente dans la paragraphe qui suit cette note, sauf qu'elle mentionnerait que dans certaines circonstances, les inférences que l'on fait quant au caractère de quelqu'un ne sont pas basées sur leurs intentions avouées, alors que dans d'autres circonstances, elles le sont. À la note 17, j'ai suggéré un mécanisme qui pourrait bien expliquer la raison de cette variation. Notons que s'ils ont raison, nous n'aurions pas un cas typique de variantisme tel que décrit par Knobe et Doris.

22 Sher défend une conception assez libérale du caractère que tous les néo-humiens ne partagent pas (voir par exemple, Smith, 2008b).

23 Pizarro et Tannenbum rapportent que les gens considèrent un individu dont le comportement est inamicale comme étant plus blâmable si son comportement est causé par le racisme que par la misanthropie et ils sont plus enclins à voir son comportement comme diagnostique de son caractère dans le premier cas.

24 Dworkin écrit que « nous avons besoin de définir en quoi consiste le fait que les motifs d'une personne soient les siens, et le fait qu'ils soient les siens propres […]. L'attitude qu'une personne adopte envers les influences qui la motivent […] détermine si oui ou non elles doivent être considérées comme « siennes ». Cette personne s'identifie-t-elle à elles, se les approprie-t-elle, se considère-t-elle comme le type de personne qui souhaite être motivée selon ces modalités spécifiques ? Dans le cas contraire, […] on ne tiendra pas ces influences pour "siennes" […] » (1976, 24). Frankfurt, parlant du cas d'un toxicomane « malgré lui », écrit dans la même veine que Dworkin « s'identifie […] à l'un […] de ses désirs opposés de premier niveau […] [il] s'approprie […] l'un d'eux plus intimement et […] prend ses distances par rapport à l'autre » (1971, p. 13). Finalement, Watson parlera d'agir selon ses propres valeurs, « c'est-à-dire ces principes et fins que [l'agent], dans des moments de calme et sans duperie de soi, considère comme étant synonyme d'une vie bonne, pleine et défendable » (105).

25 Arpaly et Schroeder s'en prennent également à cette idée : « Quand une personne se comporte de manière inhabituelle sous l'effet de l'alcool, on peut certes penser qu'elle n'est pas elle-même, mais on peut également soutenir que la chute de ses inhibitions révèle son vrai moi ; car "in vino veritas" » (1998, 261). Ces derniers proposent une autre façon de concevoir ce qu'est le véritable moi en utilisant la notion d'intégration (ou de cohérence). Je n'explorerai pas ici cette idée.

26 C'est ce que semblent également révéler d'autres études qui montrent que lorsqu'on formule les questions au sujet des attitudes explicites différemment, la dissociation est beaucoup moins prononcée. Ainsi, « il existe des données selon lesquelles les Américains de race

blanche sont conscients de leur préjugés automatiques. Alors que les étudiants de race blanche tendent à rejeter explicitement les préjugés (sous forme d'idées et de croyances) à propos des minorités, à peu près 75 % d'entre eux sont d'accord avec des énoncés du type de celui-ci, que l'on trouve dans un questionnaire : "Même si je ne le veux pas nécessairement, j'ai parfois des réactions basées sur des préjugés […] à l'égard des minorités raciales que je ne suis pas sûr de pouvoir empêcher" » (Uhlmann et Nosek, sous press).

27 La question du contrôle est donc différente de celle de la motivation dont nous avons parlé à la section précédente. On pourrait imaginer que quelqu'un qui n'a que des motivations externes à contrôler les comportements provoqués par les attitudes implicites réussisse quand même, par les moyens que je présente dans la section, à contrôler son comportement. Ce qui distinguera les deux types de motivation est la façon par laquelle ils affectent le comportement.

28 Voir Mendoza et al. (2010) qui proposent pour leur part une variation de l'expérience de Stewart et Payne, laquelle consiste, non pas à substituer une interprétation contre-stéréotypique à une interprétation stéréotypique, mais plutôt à implanter l'exécution d'une procédure à effectuer lorsque l'indice est présenté. Ils font donc répéter aux sujets des phrases comme : « Si je vois une personne de race noire, je vais répondre de façon circonspecte. » Exposés à de telles phrases, les sujets ont des résultats semblables à ceux des sujets de Stewart et Payne.

29 On dira d'un comportement qu'il est contrôlé, lorsque ce comportement est sélectionné ou rejeté, modifié ou modulé selon qu'il conduit ou non à l'atteinte d'un certain but et qu'il cesse lorsque le but en question est atteint. On distinguera donc le contrôle effectué par un but de l'effet causal pur et simple qui n'est pas guidé par un but (par exemple, une attitude implicite qui cause une réponse).

30 J'ai donné des versions de cet article en conférence au Center for Cognitive Sciences and Semantics de l'université de Riga, au département de philosophie du Cégep de Sherbrooke et de l'université Western Ontario. Je remercie ceux qui ont assisté à ces conférences pour leurs questions et leurs commentaires. J'aimerais également remercier Christine Tappolet, Edouard Machery ainsi que les deux évaluateurs de cet article pour leurs nombreux commentaires sur la version précédente de ce texte. La rédaction de cet article a été rendue possible grâce à une subvention ordinaire du CRSH.