La philosophie expérimentale et l'intuition compatibiliste/incompatibiliste

J.-B. Guillon

Extrait du chapitre 1

La plupart des auteurs incompatibilistes commencent leur présentation du problème du libre arbitre en affirmant que notre conception spontanée ou naturelle du libre arbitre non seulement « génère un conflit » avec le déterminisme, mais est même tout bonnement incompatibiliste. Par exemple, Robert Kane écrit ceci :

D'après mon expérience, la plupart des personnes ordinaires sont au départ des incompatibilistes naturels. Elles croient qu'il y a un certain type de conflit entre liberté et déterminisme ; et l'idée que la liberté et la responsabilité pourraient être compatibles leur semble tout d'abord n'être qu'une « dérobade lamentable » (quagmire of evasion) (William James) ou un « misérable subterfuge » (Emmanuel Kant). C'est seulement par les arguments subtils des philosophes qu'on peut faire sortir les personnes ordinaires de cet incompatibilisme naturel. (Kane 1999, 217)

On trouvera aisément le même genre de déclarations chez la plupart des auteurs libertariens[1]. On en trouvera également chez les auteurs indéterministes durs, notamment chez Galen Strawson, un des premiers auteurs à s'être intéressé de près au rôle et au contenu de l'expérience commune du libre arbitre :

[La conception incompatibiliste] est exactement le type de liberté que la plupart des personnes croient posséder de manière ordinaire et non-réflexive. (Strawson 1986, 30). Il fait partie de notre nature de considérer que le déterminisme pose un sérieux problème à nos notions de responsabilité et de liberté. (Strawson 1986, 76)[2].

Jusqu'à une époque récente, les déclarations de ce genre reposaient essentiellement sur l'expérience personnelle de l'auteur – soit auprès de ses étudiants, soit auprès de ses amis non-philosophes. Si un Robert Kane, par exemple, a toujours rencontré chez tous ses étudiants débutants et tous ses amis non philosophes une grande difficulté à prendre au sérieux le compatibilisme, au moins initialement, il était sans doute justifié à extrapoler son expérience et à supposer que les autres professeurs rencontraient le même genre de réactions[3]. Mais quelle que soit la valeur de cette expérience informelle, elle a fini par attirer le scepticisme de certains auteurs compatibilistes, qui ont fait observer trois biais possibles : d'une part, il est possible qu'un auteur incompatibiliste présente le problème à ses étudiants d'une manière qui incite à être

---

1 (Nahmias et al. 2006, 29) donnent une excellente revue de ce genre de déclarations. Ils citent notamment (Campbell 1951, 451), (O'Connor 2000, 4), ou encore (Ekstrom 2002, 310) : « we come to the table, nearly all of us, as pretheoretic incompatibilists ».

2 Pour le même genre déclarations chez d'autres auteurs incompatibilistes durs, voir notamment (Pereboom 2001, xvi) et (Smilansky 2003, 259).

3 On peut d'ailleurs imaginer que Robert Kane s'appuyait également sur la confirmation du témoignage de ses collègues

incompatibiliste (quand bien même il ferait des efforts sincères pour donner une présentation neutre) ; deuxièmement, il se pourrait qu'il surinterprète les réactions de ses étudiants dans le sens de sa thèse ; troisièmement, il se pourrait que ses étudiants soient effectivement incompatibilistes, mais que cela manifeste seulement un biais sociologique de tel échantillon de population, et non pas un quelconque jugement naturel. Ce genre de précautions – que les auteurs incompatibilistes trouvent généralement indûment scrupuleuses compte tenu de l'expérience pédagogique largement partagée – ont eu le grand mérite de susciter depuis les années 2000 une série d'enquêtes expérimentales pour tester l'existence et la naturalité des intuitions incompatibilistes chez les gens ordinaires.

Les études de « philosophie expérimentale » (X-Phi) sur le libre arbitre dans les dix dernières années ont essentiellement été le fruit de deux équipes de recherche, l'une constituée d'Eddy Nahmias et divers associés, l'autre de Shaun Nichols, Joshua Knobe, et divers associés. Bien que cette discussion soit encore très vive et largement ouverte, on peut déjà en tirer quelques enseignements particulièrement importants. Le plus simple est de présenter l'histoire de cette discussion en trois grandes étapes.

Les recherches de X-Phi sur le libre arbitre ont été lancées, à l'origine, par l'équipe d'Eddy Nahmias, qui avait pour objectif principal de remettre en question l'idée reçue d'un incompatibilisme spontané. Les résultats de ces premières recherches se sont avérés surprenants ; en effet, ils semblaient montrer qu'en réalité les intuitions communes étaient compatibilistes et non pas incompatibilistes. La première étude importante, (Nahmias et al. 2004), demande aux sujets interrogés de donner une interprétation de l'expression « j'aurais pu faire autrement » (parmi plusieurs interprétations proposées) et semble indiquer que, pour une majorité de sujets, l'interprétation privilégiée est celle qui correspond à la conception compatibiliste des possibilités alternatives[4]. Je me concentrerai ici sur la deuxième étude importante de Nahmias et al., à savoir (Nahmias et al. 2005). Dans cette étude, Nahmias et al. commencent par décrire aux sujets interrogés un monde déterministe (en prenant soin de le décrire d'une manière qui n'entraîne aucune confusion avec une forme de fatalisme ou de manipulation). Puis ils situent à l'intérieur de ce monde déterministe un agent qui commet une certaine action. Les sujets sont ensuite invités à dire si l'agent en question avait, dans ce monde déterministe, la liberté et la responsabilité morale pour son acte. Le scénario d'origine utilisé par Nahmias et al. était le suivant :

Scénario : Imaginez qu'au siècle prochain nous découvrions toutes les lois de la nature, et que nous construisions un superordinateur capable de déduire, à partir de ces lois de la nature et de l'état actuel des choses dans l'univers, ce qui va se produire exactement dans le monde à

[4] Les descriptions proposées pour l'expression « j'aurais pu faire autrement » étaient les suivantes : « A. 'I could have chosen to do otherwise even if everything at the moment of choice had been exactly the same'. B. 'I could have chosen to do otherwise only if something had been different (for instance, different considerations had come to mind as I deliberated or I had experienced different desires at the time)'. » (Nahmias et al. 2004, 174). L'interprétation B est ensuite présentée par Nahmias et al. comme «l'interprétation compatibiliste », et le résultat donne une approbation de cette interprétation « compatibiliste » à 62%. On peut contester l'interprétation que Nahmias et al. donnent de leur résultat de deux manières : d'une part, il est possible que les sujets considèrent les deux scénarios comme possibles et le scénario B (choix différent avec considérations différentes) comme seulement plus courant ou plus probable. Deuxièmement, il n'est pas évident que le scénario B soit réellement compatibiliste dans l'esprit des sujets, dans la mesure où les considérations elles-mêmes, qui mènent au choix, peuvent tout à fait être conçues comme étant (au moins dans une certaine mesure) au pouvoir du sujet. Dans ce cas, les gens pourraient fort bien n'avoir pas voulu nier au sujet un véritable contrôle des alternatives, mais seulement lui attribuer un contrôle indirect (antérieur) des alternatives.

n'importe quel moment futur. Il peut regarder tout ce qui se passe dans l'univers, et prédire tout ce qui lui arrivera avec une précision de 100%.

Supposez qu'un tel superordinateur existe, et qu'il regarde l'état de l'univers à un certain moment, le 25 mars 2150 ap. J.C., 20 ans avant la naissance de Jeremy Hall. L'ordinateur déduit ensuite à partir de cette information et des lois de la nature que Jeremy va dévaliser la Fidelity Bank à 18h00 le 26 janvier 2195. Comme toujours, la prédiction du superordinateur est exacte ; Jeremy dévalise la Fidelity Bank à 18h00 le 26 janvier 2195. (Nahmias et al. 2005, 566)

Les sujets devaient ensuite répondre aux deux questions suivantes :

FW : Croyez-vous que, en dévalisant la banque, Jeremy a agi « of his own free will » (de son plein gré / « par son libre arbitre ») ? MR : Croyez-vous que Jeremy est moralement blâmable pour avoir dévalisé la banque ?

A chacune de ces questions, les sujets interrogés on répondu positivement de manière très significative, en dépit du scénario déterministe : 76% de oui pour la question FW, 83% de oui pour la question MR (Nahmias et al. 2005, 566–568). La conclusion de cette étude révolutionnaire semblait sans appel : les intuitions communes sont en réalité compatibilistes ; la majorité d'entre nous peut concevoir le libre arbitre et la responsabilité même dans un scénario déterministe.

Pourtant, une deuxième étape de la discussion a renversé la conclusion au moyen d'études empiriques plus approfondies. Dans leur étude (Nichols and Knobe 2007), Shaun Nichols et Joshua Knobe ont enrichi le dispositif expérimental dans deux directions ; ils ont bien sûr retrouvé le résultat de Nahmias et al., mais ce résultat s'est trouvé placé dans un contexte général qui permet de conclure que les sujets ordinaires ont, en fait, des intuitions incompatibilistes.

La première addition apportée au dispositif de Nahmias et al. a été d'ajouter une première question avant la question de libre arbitre ou la question de responsabilité morale, une question qu'on pourrait appeler « question sur le monde actuel » ou « question d'actualité ». Après avoir décrit l'univers déterministe dans des termes semblables à ceux de Nahmias et al. (l'Univers A), Nichols et Knobe décrivent un deuxième univers, qui contient certains indéterminismes, situés au moment des décisions humaines (l'Univers B). Puis ils posent aux sujets la question suivante :

QA : Lequel de ces deux univers est, d'après vous, le plus semblable au nôtre ? (Nichols and Knobe 2007, 669)

A cette question, presque tous les sujets (plus de 90%) ont répondu que notre univers ressemblait davantage à l'Univers indéterministe B. Ce résultat a une conséquence considérable pour l'interprétation de l'expérience de Nahmias et al. Il montre en effet que les actions libres telles que les gens se les représentent effectivement, ou dans le monde actuel, supposent des indéterminismes. Et donc, le résultat de Nahmias et al. nous informe seulement sur la manière dont varient nos intuitions lorsque nous considérons un monde différent du nôtre : si nous vivions finalement dans un monde déterministe, nous garderions sans doute l'attitude d'attribuer libre arbitre et responsabilité morale. Est-ce là une intuition compatibiliste ? Non, et la meilleure façon de le voir est de constater qu'un auteur aussi incompatibiliste que Peter van Inwagen partage cette intuition : (…) van Inwagen (1983, 219) soutient que si on lui prouvait scientifiquement l'existence du déterminisme, alors il n'en conclurait pas que nous ne sommes pas libres : il en conclurait plutôt que ses arguments incompatibilistes doivent, finalement, contenir une erreur quelque part. Sa croyance au libre arbitre (et à la responsabilité morale) est plus forte que sa

croyance incompatibiliste, et l'emporterait donc dans une telle éventualité. Mais en l'état actuel des choses, van Inwagen n'en est pas moins incompatibiliste : il croit que nous sommes libres d'une manière qui introduit des indéterminismes dans la nature. Et le résultat de la question d'actualité QA semble montrer de manière très convaincante que les personnes ordinaires ont le même genre d'attitude : si on les place dans un scénario déterministe, elles continueraient à attribuer libre arbitre et responsabilité, mais (pour elles) ce scénario est purement contrefactuel, car dans le monde actuel les actions libres engagent certains indéterminismes. Pourquoi préserverions-nous ces attitudes même dans un scénario déterministe ? Le deuxième apport de Nichols et Knobe a permis d'apporter un élément de réponse à cette question.

Pour enrichir encore le dispositif expérimental de Nahmias et al., Nichols et Knobe ont fait observer que les expériences menées jusque là concernaient seulement des cas concrets (comme Jeremy dévalisant une banque) susceptibles de produire en nous des affects, et de biaiser nos jugements conceptuels sur la compatibilité du déterminisme et du libre arbitre. Nichols et Knobe ont donc enrichi le dispositif d'une question abstraite sur l'univers déterministe A :

QAbstraite : Dans l'univers A, est-il possible pour une personne d'être pleinement responsable de ses actions ? (Nichols and Knobe 2007, 670)

Or à cette question abstraite, une écrasante majorité de sujets (86%) a donné la réponse incompatibiliste[5]. Nichols et Knobe ont également posé une question concrète (comme Nahmias et al.) – le cas de Bill qui bat sa femme et ses enfants jusqu'à la mort – et ils ont retrouvé le résultat plutôt compatibiliste dans ce cas concret. Mais ce résultat semble pouvoir s'interpréter comme l'effet des émotions engagées par le cas concret : si l'on veut découvrir la vraie intuition conceptuelle, il faut mettre de côté les émotions, ce que fait la question abstraite. C'est ce que Nichols et Knobe appellent le « modèle de l'erreur de performance affective » (Nichols and Knobe 2007, 671–672) : lorsque nous appliquons nos concepts de libre arbitre ou de responsabilité sans affect, notre performance conceptuelle est bonne (nous appliquons bien ce concept) ; en revanche, lorsque nos émotions sont éveillées par la description d'un acte horrible, ces émotions biaisent l'application de nos concepts et nous poussent à attribuer la responsabilité même si une bonne performance conceptuelle requerrait de retenir cette attribution. La grande force de l'analyse de Nichols et Knobe, c'est qu'elle ne nie pas totalement l'existence des intuitions compatibilistes de sens commun : pour eux, le sens commun permet à la fois de délivrer certaines intuitions compatibilistes et certaines incompatibilistes. Mais lorsqu'on regarde les causes de cette divergence, tout laisse à penser que les secondes représentent plus fiablement notre compétence conceptuelle en tant que telle.

Pour interpréter cette divergence entre cas concret et cas abstrait, une autre interprétation a été proposée, le « modèle de la compétence concrète » (Nichols and Knobe 2007, 673–674). L'idée générale est la suivante : lorsqu'on interroge les gens sur un cas concret, on les invite à appliquer leur concept, par exemple leur concept de libre arbitre, et on obtient ainsi des jugements appliquant ce concept. En revanche, lorsqu'on pose la question abstraite, on invite les gens à réfléchir à propos de ce concept, et ce qu'on obtient alors, c'est la théorie qu'ont les gens à propos de ce concept. Or les gens pourraient fort bien se faire une fausse théorie de leurs propres

[5] Dans cette expérience, Nichols et Knobe n'ont pas posé de question abstraite à propos du libre arbitre proprement dit. Mais des études postérieures ont retrouvé le même résultat incompatibiliste en posant une telle question ; en particulier, (Roskies and Nichols 2008, 375–376) ont observé une très forte approbation à l'affirmation suivante : « it is impossible for people in Universe A to make truly free choices ».

concepts ; donc la manière la plus sûre et fiable de tester le compatibilisme ou l'incompatibilisme du concept de sens commun, c'est de proposer aux sujets des cas concrets et de leur demander leurs jugements, pas de les inviter à donner leur théorie. Ce « modèle de la compétence concrète » permet de poser une distinction très importante pour mon propos : la distinction entre « concept de Sens Commun du libre arbitre » et « conception de Sens Commun du libre arbitre ». Je reprends ici la terminologie de Dana Nelkin :

> Il sera utile de [...] distinguer entre le concept de liberté et les conceptions particulières de la liberté. Premièrement, il y a un concept de liberté que tout le monde (ou en tout cas beaucoup de gens) possèdent, même s'ils sont en désaccord sur les conditions nécessaires et suffisantes de son instanciation. Il doit y avoir un tel concept si l'on veut pouvoir dire que les compatibilistes, incompatibilistes, et ceux qui croient la liberté impossible, sont en désaccord. Bien sûr, il est possible de défendre que tout le problème du débat sur le libre arbitre est que les participants parlent en fait différents langages [...] Mais je crois que cette idée fait trop peu de crédit aux participants de la discussion. [...] Deuxièmement, il y a aussi les différentes analyses de la liberté proposées [...] Ce sont des conceptions de la liberté, des tentatives pour formuler les conditions dans lesquelles les actions sont libres. (Nelkin 2004, 113–114)

Le « concept [...] que tout le monde possède », dont parle Nelkin, est clairement le concept de Sens Commun que j'essaye de définir de manière descriptive dans ce chapitre, c'est-à-dire le référent commun (quoique non analysé) des différentes discussions sur le libre arbitre. Quant aux conceptions du libre arbitre, Nelkin pense ici principalement aux conceptions philosophiques ; mais comme on vient de le voir, il se pourrait aussi que les gens ordinaires aient une tendance naturelle à se former une certaine conception du libre arbitre – une conception qui pourrait être systématiquement erronée, c'est-à-dire infidèle à la définition véritable du concept qu'ils tentent d'appliquer. Par conséquent, lorsqu'on se demande si le Sens Commun est compatibiliste ou incompatibiliste, on peut vouloir poser deux questions différentes :

(Question du concept de Sens Commun) : le concept commun de « libre arbitre » est-il logiquement compatible avec le concept de déterminisme ?

(Question de la conception de Sens Commun) : selon la conception que nous nous faisons naturellement du « libre arbitre », celui-ci est-il compatible avec le déterminisme ?

Il se pourrait très bien que « le Sens Commun » soit incompatibiliste dans le sens de la question 2, tout en ayant un concept compatibiliste dans le sens de la question 1. J'ai déjà évoqué ce genre de possibilité lorsque j'ai montré que la définition descriptive du libre arbitre par le conflit intuitif ne préjugeait pas de la question de compatibilité : si elle n'en préjuge pas, c'est justement parce qu'une conception incompatibiliste (un conflit intuitif) peut reposer sur un concept qui, en réalité, est compatibiliste.

La conclusion à tirer (…) c'est que le « modèle de la compétence concrète » ne remet pas en cause l'existence du conflit intuitif que nous recherchons (…). Ce que nous recherchons ici, ce n'est pas une preuve que le concept de Sens Commun est effectivement incompatibiliste, mais seulement qu'il a une propension à générer un conflit intuitif. Cette propension pourrait être liée à une incompatibilité réelle entre les concepts de libre arbitre et de déterminisme (ainsi que le défendent Nichols et Knobe) ou bien à une erreur systématique de théorisation sur ce concept (ainsi que le défend le modèle de la compétence concrète). Nichols et Knobe ont apporté des réponses expérimentales solides pour rejeter le modèle de la compétence concrète et sa

conclusion compatibiliste[6]. (…) Ce qui nous intéresse ici, c'est que Nichols et Knobe ont établi de manière incontestable (et incontestée) les deux résultats suivants : 1. la conception que nous nous faisons spontanément de nos actions est qu'elle introduit des indéterminismes dans l'univers (c'est l'apport de la « question d'actualité ») 2. nous avons une tendance naturelle à concevoir (au niveau abstrait) le libre arbitre comme incompatible avec le déterminisme (c'est l'apport de la « question abstraite »)

Aucune des discussions qui ont suivi dans la littérature de X-Phi n'ont permis de mettre en cause significativement ces deux résultats[7]. Et quelle que soit l'importance qu'on accorde aux intuitions « compatibilistes » exprimées sur les cas concrets, elles n'altèrent en rien ces deux faits. Or ces deux faits sont suffisants pour établir qu'il y a, bel et bien, une tendance naturelle à expérimenter un certain conflit intuitif entre l'image d'un monde déterministe et l'image que nous nous faisons naturellement de nous-mêmes et de nos actions libres.

Pour être plus exact, il a fallu attendre la troisième phase d'études, et en particulier (Sarkissian et al. 2010), pour arriver véritablement à la conclusion d'un conflit naturel et non pas seulement culturel. En effet, les premières études ont toutes été menées sur des sujets américains. Leur résultat aurait donc pu manifester un biais de sélection sociologique. Sarkissian et al. ont donc réalisé une étude plus large, à partir de quatre échantillons d'étudiants : 66 étudiants des Etats-Unis, 55 étudiants indiens, 40 de Hong Kong, et 70 de Colombie. A ces étudiants, ils ont posé la question d'actualité et la question abstraite de compatibilité (donc les deux questions qui délivraient un résultat incompatibiliste dans l'étude de Nichols et Knobe). Le résultat est très clair : dans chacun des échantillons, on retrouve le résultat très majoritairement incompatibiliste, aussi bien pour la question d'actualité (notre univers ressemble plus à l'univers B) et pour la question abstraite (dans l'univers A, il n'est pas possible pour une personne d'être pleinement responsable moralement de ses actions). Il y semble donc bien y avoir un conflit intuitif universel (inter-culturel) entre l'image d'un univers déterministe et notre conception de nous-mêmes et de nos actions.

---

[6] La stratégie employée dans (Nichols and Knobe 2007, sec. 6 sqq) a consisté à distinguer le caractère concret du scénario et sa dimension affective. La question sous-jacente était la suivante : la différence de résultat entre question concrète et question abstraite est-elle due au potentiel affectif de la question concrète (comme le prédit le modèle de l'erreur de performance affective) ou bien spécifiquement au caractère concret de la question (comme le prédit le modèle de la compétence concrète). Pour répondre à cette question, Nichols et Knobe ont posé deux questions concrètes, l'une à fort potentiel affectif (Bill stalks and rapes a stranger), l'autre sans potentiel affectif (Mark arranges to cheat on his taxes). L'expérimentation a révélé qu'à la question concrète sans potentiel affectif, les sujets avaient une très forte tendance à répondre comme à la question abstraite (pas de responsabilité morale), et que c'est seulement la question concrète à fort potentiel affectif qui faisait apparaître le résultat compatibiliste. Autrement dit, l'intuition compatibiliste ne vient pas du caractère concret ou non de la question envisagée, mais bel et bien d'un investissement affectif, comme le prédit le modèle de l'erreur de performance affective.

[7] Les objections qui ont été faites contre le résultat de Nichols et Knobe – outre l'objection « jugement » vs « théorie » qu'on vient de voir – avaient toutes à voir avec la description des univers ou la formulation des scénarios. La description de l'univers déterministe aurait employé des termes qui portaient à une confusion avec le fatalisme ; ou la question de responsabilité morale était posée en des termes trop forts. Le principal problème de ce genre d'objections, c'est que Nichols et Knobe ont utilisé les mêmes formulations pour la question concrète et pour la question abstraite, et que dans le cas concret, ces formulations ne semblent pas avoir eu d'effet de biais puisque le résultat « compatibiliste » de Nahmias et al. a pu être retrouvé. Donc le résultat incompatibiliste du cas abstrait (et surtout l'écart entre les deux) ne peut pas être attribué à un problème de formulation. Pour répondre à ces problèmes de formulations, plusieurs études ont par ailleurs été effectuées avec des formulations non controversées, et les mêmes résultats ont été trouvés. Pour un résumé de ces discussions et études, voir notamment (Sarkissian et al. 2010, 349).

Extrait du ch. 4

Concernant la confusion entre déterminisme et fatalisme ou mécanisme, certains résultats récents en philosophie expérimentale seront particulièrement pertinents. En effet, dans une série d'études, en particulier (Nahmias, Coates, and Kvaran 2007; Nahmias and Murray 2011), Eddy Nahmias s'est attaché à montrer que nous avions une tendance naturelle à confondre le déterminisme avec d'autres théories plus radicales (en particulier le modèle de court-circuit), et que cette confusion expliquait nos intuitions incompatibilistes. Pour montrer cela, (Nahmias and Murray 2011) ont cherché à évaluer deux choses, d'une part notre tendance à mal comprendre le déterminisme, et d'autre part une éventuelle corrélation entre de telles mécompréhensions et l'intuition incompatibiliste telle qu'on la trouve dans les expériences de X-Phi. Nahmias et Murray ont réutilisé pour cette expérience les scénarios que nous avons décrits au chapitre 1 (p.52), notamment la distinction entre Univers A (déterministe) et Univers B (indéterministe aux endroits des décisions humaines) et la distinction entre scénarios concrets et scénarios abstraits. Après avoir présenté ces scénarios, Nahmias et Murray ont posé deux séries de questions, une série classique concernant la responsabilité morale, le libre arbitre, ou la légitimité du blâme dans ces univers, et une deuxième série de question destinée à tester si les sujets interprétaient la situation considérée comme une situation de court-circuit. Voici la série de question (Nahmias and Murray 2011, sec. III) :

Décisions : Dans l'univers [A/B], les décisions d'une personne n'ont aucun effet sur ce qu'elle est finalement causalement amenée à faire.

Volontés : Dans l'univers [A/B], ce qu'une personne veut n'a aucun effet sur ce qu'elle est finalement causalement amenée à faire.

Croyances : Dans l'univers [A/B], ce qu'une personne croit n'a aucun effet sur ce qu'elle est finalement causalement amenée à faire.

Pas de contrôle : Dans l'univers [A/B], une personne n'a aucun contrôle sur ce qu'elle fait.

Passé différent : Dans l'univers A, tout ce qui arrive doit arriver, même si ce qui était arrivé dans le passé était différent.

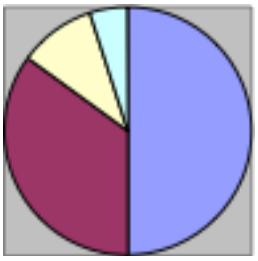
Il est clair que rien dans le déterminisme n'empêche que les décisions, les volontés, et les croyances des agents aient un effet causal. Ce qui exclut un tel effet, c'est le modèle du court-circuit, pas le modèle déterministe. Par conséquent, une réponse positive à l'une de ces questions manifestera une confusion entre déterminisme et court-circuit. De même, il est clair que dans un scénario déterministe, un passé différent entraîne un futur différent ; c'est dans un modèle fataliste que le futur est fixé quel que soit le passé. Donc une réponse positive à la dernière question manifesterait une confusion entre déterminisme et fatalisme[8].

Nahmias et Murray ont mis en évidence le résultat qu'ils souhaitaient trouver, à savoir une forte corrélation entre illusion de court-circuit et intuition incompatibiliste. Plus précisément, si l'on se concentre sur le scénario « abstrait » de Nichols et Knobe, celui dont nous avions vu qu'il révélait la plus forte intuition incompatibiliste, Nahmias et Murray ont retrouvé bien sûr cette

[8] La quatrième question est un peu plus litigieuse car la notion de contrôle n'est pas absolument claire ; mais dans la mesure où un modèle déterministe permet que nos volontés aient un effet causal sur nos actions, on peut dire qu'il permet sans doute une certaine forme de contrôle ; dans ce cas, quelqu'un qui refuserait à l'agent toute forme de contrôle dans le scénario déterministe l'aurait sans doute confondu avec un scénario plus radical, fataliste ou de court-circuit.

intuition incompatibiliste (73% de réponses négatives à la première série de questions, concernant la responsabilité morale, le libre arbitre et le blâme dans l'Univers A), mais ils ont également trouvé un taux très fort de confusion du court-circuit (66% de réponses positives – et donc erronées – à la deuxième série de questions). Ce résultat les amène à conclure que sur les 73% d'intuitions « incompatibilistes », il y en en fait la quasi-totalité (66%) qui sont de faux incompatibilistes, c'est-à-dire des personnes qui croient le déterminisme incompatible avec le libre arbitre simplement parce qu'ils comprennent mal le déterminisme. Ainsi, la forte intuition incompatibiliste que nous observons en philosophie expérimentale serait simplement fondée sur une confusion du court-circuit.

Ce résultat de Nahmias et Murray est certainement l'argument le plus fort dont on dispose aujourd'hui en faveur d'une théorie de l'illusion. L'argument n'est pourtant pas concluant pour deux raisons. Premièrement, la conclusion de Nahmias et Murray semble reposer sur la confusion entre deux probabilités conditionnelles. Appelons « compétents » les sujets qui comprennent bien le déterminisme, et « incompétents » ceux qui le confondent avec le modèle du court-circuit. Ce que Nahmias et Murray observent, c'est qu'une large partie des incompatibilistes sont des « incompétents », autrement dit, qu'on a de fortes chances d'être un « incompétent » lorsqu'on est un incompatibiliste[9]. La probabilité P (incompétent | incompatibiliste) est élevée (0,91). Mais les intuitions des incompétents, dans la mesure où ils ne comprennent pas bien le déterministe, n'est en fait ni une intuition incompatibiliste, ni une intuition compatibiliste ; c'est une intuition « hors-sujet ». Pour mettre en évidence nos intuitions compatibilistes ou incompatibilistes, il faut donc mettre de côté les incompétents. La probabilité qui nous intéresserait vraiment pour connaître si nous avons une intuition incompatibiliste ou non, c'est la probabilité d'être incompatibiliste lorsqu'on est compétent. Autrement dit P (incompatibiliste | compétent) et non pas P (incompétent | incompatibiliste). Or ces deux probabilités sont indépendantes l'une de l'autre : il se pourrait très bien que la plupart des incompatibilistes soient incompétents, et que pourtant, parmi les compétents, il reste une majorité d'incompatibilistes. Ce serait le cas, par exemple, dans le schéma suivant :



- 5 incompatibilistes incompétents
- 10 incompatibilistes compétents
- 35 compatibilistes compétents
- 50 compatibilistes incompétents

Dans une telle situation, bien que la majorité des incompatibilistes soient incompétents (50/85=59%), on ne pourrait pas conclure que l'intuition majoritairement incompatibiliste (85%) « est due à » la mécompréhension du déterminisme, puisqu'on retrouverait une forte majorité incompatibiliste en se concentrant sur les sujets compétents (35/45=77%). La vraie question, pour

[9] Ils montrent également que la proportion d'incompétents parmi les compatibilistes est beaucoup plus faible.

connaître nos intuitions communes sur le compatibilisme ou l'incompatibilisme, consiste donc bien à se demander si la majorité incompatibiliste existe toujours lorsqu'on se concentre sur les sujets compétents. C'est pour cette raison que (Deery, Bedke, and Nichols 2013, 133–134) ont mis en place dans leur dispositif expérimental une étape « d'entraînement au déterminisme » qui a pour but de s'assurer que les sujets testés comprennent bien la notion de déterminisme et ne la confondent ni avec le fatalisme ni avec le court-circuit[10].332 Le résultat de Deery et al. est très clair : lorsqu'on se concentre sur les sujets compétents, sur les sujets qui ne font pas la confusion entre déterminisme et court- circuit, on retrouve tout de même le fort taux d'intuitions incompatibilistes. La confusion entre déterminisme et court-circuit n'est donc pas « ce qui explique » la présence d'une forte intuition incompatibiliste.

Une deuxième considération peut être avancée pour répondre à Nahmias et Murray. Malgré notre première réponse (le fait que « parmi les compétents », il reste une majorité d'incompatibilistes), il n'en demeure pas moins que Nahmias et Murray ont mis en évidence une forte proportion de confusion dans la population générale. Par conséquent, il semble toujours pertinent de dire que pour la population générale au moins, une grande partie de l'incompatibiliste est fondée sur une confusion entre déterminisme et court-circuit. Mais cela n'est pas évident non plus : une corrélation statistique n'indique pas l'ordre causal ou l'ordre d'explication. Nahmias et Murray expliquent la corrélation entre incompétents et incompatibilistes en disant que c'est parce qu'on est incompétent qu'on est incompatibiliste. Cet ordre d'explication est clairement possible, mais ce n'est pas le seul : il se pourrait en effet que certains sujets soient au contraire « incompétents » parce qu'ils sont incompatibilistes. Qu'est-ce que cela voudrait dire ? Ainsi qu'on vient de l'observer, il y a une forte tendance à être vraiment incompatibiliste (à être incompatibiliste à propos du concept bien compris de déterminisme). Quelqu'un qui est un vrai incompatibiliste jugera donc que, dans l'univers déterministe A, nous n'avons ni libre arbitre, ni responsabilité morale, ni fondement au blâme etc. Toutes ces conséquences sont généralement tenues pour moralement « déprimantes ». Lorsqu'on pose ensuite à ce sujet la deuxième série de questions (dans cet univers, nos volontés ou décisions ont-elles un effet causal ? etc.), il est très possible que les conséquences déprimantes de la première réponse aient un effet de « contagion » et que le sujet réponde pour cela négativement. Autrement dit, le sujet répondrait qu'on est comme dans un univers de court-circuit non pas parce qu'il a mal compris la description de l'univers déterministe, mais parce que l'univers déterministe et l'univers de court-circuit sont équivalents du point de vue des conséquences morales qu'on vient de lui faire envisager. Des effets de contagion bien plus spectaculaires ont été observés en philosophie expérimentale. Par exemple, dans un scénario célèbre de Joshua Knobe, un chef d'entreprise choisit de mettre en œuvre un programme dont on lui a annoncé qu'il serait positif pour l'environnement, mais le chef d'entreprise ne se soucie pas le moins du monde de ces conséquences positives collatérales ; son choix est entièrement guidé par son intention de faire progresser son entreprise. Dans un second scénario comparatif, le même chef d'entreprise choisit

[10] Le dispositif est le suivant : après avoir lu une description de ce que signifie le déterminisme (appelé en l'occurrence « complétude causale » pour éviter les éventuelles incompréhensions liées au mot « déterminisme »), les sujets doivent répondre à une question dont la réponse manifeste une éventuelle confusion avec le court-circuit. Ceux qui répondent mal à cette question reçoivent une explication de leur erreur, et une deuxième chance. S'ils se trompent une deuxième fois, ils sont tout simplement éliminés de l'étude et ne reçoivent même pas les questions sur le libre arbitre. Les sujets qui réussissent l'entraînement (au premier ou au deuxième coup), sont jugés compétents pour pouvoir exprimer une intuition vraiment compatibiliste ou vraiment incompatibiliste.

un programme dont on lui a fait savoir qu'il nuirait à l'environnement ; mais là encore, le chef d'entreprise ne se soucie que de son entreprise. Dans l'expérience d'origine, ces scénarios servaient à tester les intuitions communes sur l'action volontaire : une majorité de gens répondaient que le chef d'entreprise n'avait pas volontairement aidé l'environnement, mais qu'il avait nui volontairement. L'asymétrie entre les deux réponses était l'élément intéressant., compte tenu de l'absence d'asymétrie dans l'attitude du chef d'entreprise. Dans certaines expériences plus tardives, Knobe a testé d'autres intuitions, et notamment l'intuition concernant le mot « savoir » : le chef d'entreprise savait-il qu'il aiderait/nuirait à l'environnement ? Et étonnamment, la même asymétrie est apparue dans les réponses : bien qu'il soit stipulé dans le scénario que le chef d'entreprise était au courant, un certain nombre de personnes ont eu tendance à dire que le chef d'entreprise ne savait pas qu'il aiderait l'environnement. Ceci s'explique par effet de contagion : lorsqu'on a à l'esprit l'enjeu moral de la question, le chef d'entreprise est aussi peu louable que s'il n'avait absolument rien su des effets sur l'environnement. Les deux situations sont équivalentes. De la même façon, si quelqu'un est un incompatibiliste véritable et qu'on attire d'abord son attention sur les conséquences moralement « déprimantes » de cet incompatibilisme, alors il aura des chances de ne pas porter attention à la différence entre déterminisme et court-circuit parce que cette différence n'est pas pertinente dans le contexte. Dans ce cas, la forte proportion d'intuitions de «confusion» dans la population générale ne s'expliquerait pas par le fait que de nombreuses personnes sont incompétentes et confondent effectivement déterminisme et court- circuit, mais par le fait qu'un grand nombre de personnes jugent non pertinente la distinction (qu'ils peuvent faire) entre déterminisme et court-circuit.

La leçon générale concernant la théorie de l'illusion par confusion avec le fatalisme et le court-circuit est donc la suivante : tout d'abord l'intuition incompatibiliste reste majoritaire parmi les sujets dont on s'assure qu'ils ne font aucune confusion, et d'autre part, si la confusion semble répandue dans la population générale, ce pourrait très bien être parce que la population générale est constituée (essentiellement) d'incompatibilistes compétents qui jugent non pertinente la distinction existante entre déterminisme et court-circuit.

Nahmias, Eddy, D. Justin Coates, and Trevor Kvaran. 2007. «Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions ». Midwest Studies in Philosophy 31 (1): 214–242.

Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. 2004. « The Phenomenology of Free Will ». Journal of Consciousness Studies, 11 7 (8): 162–179.

———. 2005. «Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility ». Philosophical Psychology 18 (5): 561–584.

———. 2006. « Is Incompatibilism Intuitive? ». Philosophy and Phenomenological Research 73 (1): 28–53.

Nahmias, Eddy, and D. Murray. 2011. « Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions ». New Waves in Philosophy of Action: 189–217.

Nichols, Shaun. 2004. « The Folk Psychology of Free Will: Fits and Starts ». Mind and Language 19 (5) (November): 473–502.

Nichols, Shaun, and Joshua Knobe. 2007. « Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions ». Nous 41 (4): 663–685.

Sarkissian, Hagop, Amita Chatterjee, Felipe De Brigard, Joshua Knobe, Shaun Nichols, and Smita Sirker. 2010. « Is Belief in Free Will a Cultural Universal? ». Mind & Language 25 (3) (June 1): 346–358.