

« Précis de *La nature de la volonté* »

Joëlle Proust

Philosophiques, vol. 35, n° 1, 2008, p. 109-117.

Pour citer ce document, utiliser l'adresse suivante :

<http://id.erudit.org/iderudit/018238ar>

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <http://www.erudit.org/apropos/utilisation.html>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : erudit@umontreal.ca

Disputatio

Précis de *La nature de la volonté*

JOËLLE PROUST

Institut Jean-Nicod (CNRS, EHESS, ENS)

Le propos de l'ouvrage est à la fois introductif et constructif. Il vise à exposer les deux grandes théories causales en matière d'action, celles de Donald Davidson et de John Searle, de défendre une conception alternative, la théorie volitionniste, et d'en examiner les conséquences sur la liberté de l'agir et sur l'identité personnelle. La défense de la théorie volitionniste passe par l'examen des problèmes que doivent surmonter toutes les tentatives de définir l'action qui font appel à sa structure causale-représentationnelle, en montrant que la conception volitionniste est mieux équipée pour leur apporter une réponse.

Le premier est le problème de Malebranche. Comment expliquer que nous puissions remuer notre bras à volonté, lors même que nous ne savons pas ce qui fait que nous pouvons le remuer ? Le problème de Malebranche, s'il était sans solution, donnerait un argument de poids aux philosophes qui, à la suite de Wittgenstein, rejettent la pertinence du registre causal en matière d'action. Malebranche tentait de comprendre le lien entre volonté et exécution comme une relation causale, alors qu'en fait ce lien est d'après eux rationnel : il relève non de l'efficacité causale mais de la justification. La solution du problème de Malebranche passe par plusieurs étapes. On doit la première à Donald Davidson. Elle consiste à montrer que la représentation des raisons d'agir peut être associée à l'efficacité causale dans la mesure où un même événement a des propriétés physiques et mentales. Quand on explique une action par ses raisons, on s'intéresse à l'un des aspects de l'événement mental-cérébral considéré ; même si ce n'est pas cet aspect-là qui permet de parvenir à des lois causales strictes, c'est le même événement, pris sous un autre aspect, qui intervient à titre de cause ou d'effet¹. La seconde avancée consiste à réduire l'écart représentationnel entre la représentation mentale de l'action et de son effectuation. On peut en effet montrer que l'agent peut former de ses façons d'agir une représentation plus précise que la simple formulation conceptuelle du résultat ne permet de le faire. C'est là la fonction de ce que John Searle appelle « l'intention en action ». Le contenu d'une intention en action s'articule de la manière suivante : « Cette intention en action présente tel mouvement à exécuter, et en vertu de cette présentation, elle cause tel mouvement correspondant. » Grâce au concept d'intention en action, l'effectuation du mouvement devient officiellement partie du contenu intentionnel. La troisième étape consiste à permettre au contenu de l'intention en action d'être non conceptuel, ce que suggère déjà le recours aux démonstratifs pour caractériser les contenus

1. On trouvera plus bas, dans la réponse à Daniel Laurier, une discussion de l'épiphénoménalisme des états mentaux à laquelle cette position est associée.

d'intentions en action. Searle utilise le terme de « présentation », parce que l'intention en action donne directement accès à son contenu dans une expérience caractéristique. Mais ce contenu est d'après lui conceptuel. Or on a objecté depuis que le grain « conceptuel » n'est pas adapté à la représentation de contenus « analogiques » (comme le sont les contenus perceptifs, émotionnels ou agentifs)². On verra plus loin que la théorie de la volition permet d'apporter une pierre de plus à la solution du problème de Malebranche, en exposant la structure ontologique de l'action : cette structure est circulaire et non pas linéaire, ce qui permet à la fois d'expliquer pourquoi le sujet peut ignorer les mailons corporels de son action, et de clarifier le statut de la réflexivité dans l'agir.

Le second problème des théories causales de l'agir est celui de la causalité déviante : il peut arriver que le but d'une action soit atteint en vertu de l'impact causal de l'intention de l'agent sur l'environnement, mais d'une manière qui ne permet plus de dire que l'action était intentionnelle. On connaît l'exemple de Chisholm : un neveu tente de tuer à la carabine l'oncle dont il espère l'héritage, le rate, mais déclenche la fuite d'une harde de sangliers, à l'occasion de laquelle l'oncle est tué. Quoique l'oncle soit mort, et qu'il le soit parce que le neveu a pressé sur la gâchette, l'action finale ne satisfait pas les conditions de l'action, parce que les moyens d'agir ne sont pas ceux que l'agent avait en tête. Cet exemple n'est en rien exceptionnel. Il y a une variété de cas de causalité déviante selon qu'elle affecte tel ou tel niveau de la séquence menant des raisons d'agir au résultat final de l'action. C'est du fait de ce problème que Davidson renonce à définir l'action intentionnelle : il parvient à une condition nécessaire, mais non suffisante de l'action intentionnelle. Le concept searlien d'intention en action, dont on a vu qu'il précisait le « comment » de l'action, vise explicitement à répondre à ce problème. Il permet d'écarter les cas de déviance présentés ci-dessus, en indiquant « la condition de satisfaction » de l'intention en action : la victime doit mourir par balle du fait du tir à la carabine. Mais cette stratégie, Searle le reconnaît, ne réussit pas à bloquer entièrement la menace de la déviance causale. En effet, le lien entre l'intention en action et le mouvement corporel peut lui aussi être manipulé par un malin génie. Si, chaque fois que j'ai l'intention de lever le bras, le signal est envoyé sans que je le sache à un autre agent qui peut relayer ou non le signal vers le membre, l'action réussie ne sera pas la mienne, quoique mon intention ait pu causer le mouvement. La structure causale déviante s'oppose de nouveau à ce que le lien causal de l'intention en action avec le mouvement corporel puisse *définir* ce qu'est une action.

Searle propose d'exclure ce type de cas par convention, en ajoutant que « la relation de l'intention en action à ses conditions de satisfaction exclut l'intervention d'autres agents ou d'autres états intentionnels » (1985, 138). La

2. C'est pour cela que plusieurs auteurs, suivant les suggestions de Peacocke et d'Evans relatives au contenu non conceptuel de la perception, ont avancé que le contenu des intentions en action devait lui aussi inclure des représentations non conceptuelles, impliquant entre autres des référentiels égocentriques ou allocentriques, des propriétés de rythme et de fluidité, etc

solution nominale est-elle le seul recours ? Là aussi, la théorie de la volition permet, on le verra, de faire un pas de plus, en montrant que même les cas où l'action est parasitée par un agent extérieur, il existe un critère non nominal permettant de dire quand l'action est intentionnelle. Quels que soient les mérites de la solution apportée, on ne doit pas méconnaître qu'existe nécessairement une limite asymptotique aux manières de prémunir un organisme contre la causalité déviante : elle demeure nécessairement possible, du seul fait que nous identifions des mécanismes causaux qui, en tant que tels, peuvent toujours être perturbés. Une théorie qui prémunirait une théorie causale contre *toute* forme de déviance serait ainsi en contradiction avec ses propres présupposés³. Cette remarque ne dispense pas de chercher à donner une définition qui soit à la fois nécessaire et suffisante de l'action, en montrant que les cas de déviance sont ceux que laisse prévoir la théorie proposée.

Le troisième problème est le problème de Wittgenstein : « Que reste-t-il si l'on soustrait le fait que mon bras se lève du fait que je lève le bras ? ». En d'autres termes, qu'est-ce qui distingue mon action de lever le bras du résultat auquel elle aboutit, à savoir que mon bras soit levé ? Étant donné qu'une action reste une action quand elle se résume à une tentative inaboutie, la question de Wittgenstein doit recevoir une réponse claire. Les théories de l'action de Davidson et de Searle ne peuvent pas identifier ce « reste » parce qu'en tant que théories des intentions, elles s'exposent au cas fréquent des illusions intentionnelles. Dans les deux cas, les théories avancées conduisent à considérer comme actions des cas où aucune action n'a, en réalité, été accomplie ; il suffit que l'agent *croie* qu'il peut agir, qu'il ait des *raisons* de le faire, et qu'il *observe le résultat attendu en conséquence de son intention* (ou de sa raison d'agir) pour qu'il y ait eu action. Mais cette lacune théorique provient du fait que ces théories traitent bien des intentions d'agir et des croyances nécessaires à l'action, mais ne disent rien du déclenchement ni du contrôle de l'action. Or il est facile de voir, sur un exemple, que les conditions de type épistémique ou conatif ne sont pas suffisantes. Hugh McCann⁴ prend l'exemple d'un apprenti yogi qui croit à tort qu'il peut accélérer ses propres battements cardiaques. L'état involontaire d'excitation que crée chez lui l'idée d'y parvenir lui fait battre le cœur plus vite. Soustrayons l'effet. Reste-t-il une action ? Davidson doit l'affirmer, puisque le sujet a une raison d'agir : il croit pouvoir le faire, il désire le faire, et il obtient ce résultat parce qu'il a cette croyance et ce désir. Même si l'action échouait, elle resterait une action du fait de sa structure. Searle doit aussi admettre que l'exemple répond aux conditions de

3. Le problème de la déviance causale ne constitue une difficulté que dans les cas où il s'agit d'exposer la structure causale des représentations du contenu de l'action intentionnelle réussie. Mais il ne porte pas sur la limite de fiabilité de mécanismes qui, on le constate dans de multiples cas, ne sont pas à l'abri d'interférences ou de perturbations pathologiques. L'existence de perturbations dans l'efficience des représentations de l'agir intentionnel est au contraire une prédiction pour toute théorie causale représentationnelle de l'agir et une marque de leur adéquation.

4. McCann (1974, 460-1).

l'action : le sujet a l'intention d'accélérer ses battements cardiaques, et cette intention préalable cause l'expérience correspondante (l'intention en action, ici illusoire). A ce problème, le volitionniste s'efforce d'apporter une réponse différente : non, l'apprenti yogi n'a pas exécuté d'action, parce qu'il ne peut pas effectivement contrôler son battement de cœur ; ne disposant pas de ce savoir-faire, il ne peut pas lancer l'action correspondante.

Mais voyons maintenant en quoi consiste au juste la théorie de la volition. Toute théorie volitionniste part de l'hypothèse qu'on ne peut agir sans former une volition. Ce qui est essentiel dans l'agir, c'est non pas l'arrière-plan rationnel explicite qui conduit le sujet à agir et qui justifie l'action ; non pas l'intention que le sujet a pu former avant d'agir, mais l'opération volontaire qui préside à l'exécution de l'action. Comme l'observe Locke :

Une chose du moins qui est évidente, à mon avis, c'est que nous trouvons en nous-mêmes la puissance de commencer ou de ne pas commencer, de continuer ou de terminer plusieurs actions de notre esprit, et plusieurs mouvements de notre corps, et cela simplement par une pensée ou un choix de notre esprit, qui détermine et commande, pour ainsi dire, que telle ou telle action particulière soit faite, ou ne soit pas faite (J. Locke, 1690, II, XX1, 5, 182).

Deux objections sont fréquemment opposées à ce type de définition. Si l'on considère que la volition est elle-même une action mentale présidant au déroulement du mouvement corporel et de ses effets, on rend le concept d'action indéfinissable puisqu'il reparait dans la définition. Pire encore : cette action mentale semble nécessiter à son tour une volition, et ainsi de suite à l'infini. En fait, on peut dissoudre ces deux objections en examinant de plus près la structure réflexive de la volition. La volition est l'effort de réaliser un changement en vertu de cet effort. Cet effort est constitué par les modifications du sujet qui se met en mesure de produire l'effet visé. Pour bien comprendre pourquoi ces deux dernières formules ne présentent pas de risque de régression à l'infini ni de circularité, il faut analyser plus en détail ce qu'est un épisode de contrôle volontaire et mettre en évidence la structure téléologique de la volition. Appelons « représentation exécutive » le modèle dynamique d'un mouvement (ou d'une activation mentale) et de son bénéfice. L'ensemble de ces représentations exécutives constituent le répertoire dans lequel les volitions sélectionnent leur contenu. Comment cette sélection procède-t-elle ?

La sélection d'un contenu volitif dépend de ce que les théoriciens ont appelé « loi de l'effet » ou « principe de l'action-effet ». Ce principe téléologique pose que l'agent a tendance à sélectionner les actions qui, dans le passé, ont eu un effet favorable, et à écarter celles qui ont eu un effet défavorable. Le renforcement étant indépendant du contenu particulier de l'action, il s'opère même lorsque le sujet est incapable d'analyser les divers facteurs causalement impliqués dans la réussite de l'action. Pour illustrer cette remarquable propriété, Marteniuk et al. (1987) ont montré, par exemple, que lorsque des sujets doivent saisir un disque soit pour l'insérer dans une fente, soit pour le lancer,

la cinématique de leur geste est d'emblée différente. Les conséquences attendues de l'action modulent le début du mouvement. Seule l'analyse téléologique permet d'expliquer ce curieux phénomène de préadaptation du mouvement, qui échappe d'ailleurs à la conscience de l'agent. Ainsi, par parenthèse, le problème de Malebranche reçoit-il enfin une réponse précise. Si le sujet n'a pas à connaître ce qui lui permet d'exécuter son action, c'est en vertu du fait que la volition a une structure circulaire régie par le principe de l'action-effet (nous reviendrons plus loin, dans la réponse à Daniel Laurier, sur les raisons métaphysiques de voir dans le principe d'action-effet le guide heuristique de la définition téléologique de la volition).

Quelque éclairant qu'il soit, le principe de l'action-effet ne peut suffire à rendre compte de la complexité des actions qu'un sujet peut accomplir. S'il explique pourquoi un type d'action est sélectionné et reproduit, il n'explique pas encore comment le mouvement peut être chaque fois *ajusté et modifié* pour répondre aux contraintes éminemment variables auxquelles l'agent doit s'adapter. La généralisation du principe téléologique exige que l'on rappelle en quoi consiste le contrôle adaptatif, c'est-à-dire la forme de contrôle qui opère dans des contextes partiellement inconnus. Pour interagir avec un environnement incertain, l'organisme doit pouvoir utiliser les rétroactions antérieures pour sélectionner une représentation exécutive particulière étant donné un but nouveau, la corriger, et vérifier l'atteinte du but. Il est important, dans ce cas, de prédire intérieurement les effets qui vont se dérouler durant l'exécution, et comparer avec ces prédictions les observations recueillies. On l'aura compris, ce sont les représentations exécutives qui permettent une telle prédiction intérieure. L'élément capital du contrôle adaptatif est ainsi le *comparateur*, qui permet d'évaluer et de corriger l'action avant même qu'elle ne soit exécutée, sur la base du feedback interne de la volition.

Les deux types de contraintes qui président à la sélection volitive et à l'opération du comparateur sont bien connues par les mathématiciens du contrôle adaptatif (Aubin, à paraître). 1) Les « lois de régulation » associent à chaque nouveau contexte les commandes susceptibles d'atteindre une cible étant donné les contraintes. 2) les « lois de rétroaction » déterminent quelle partie de l'espace de régulation l'organisme peut atteindre à une étape donnée de son apprentissage. Confrontés à ces deux types de contraintes, les agents mémorisent les solutions efficaces et construisent des « modèles internes » des divers contextes exécutifs⁵.

Pour résumer ce que l'on vient de survoler trop rapidement, un épisode volitif consiste à activer un ensemble hiérarchisé de représentations exécutives. La volition de faire X en vue de P (tendre la main pour saisir un verre) dépend alors nécessairement de trois conditions :

5. On trouvera plus loin, dans la réponse à Pierre Livet, une analyse de la pertinence de la théorie de la viabilité pour comprendre l'action.

- 1) Contrôle existant : il existe dans l'espace de régulation accessible à l'agent pour ce contexte au moins une trajectoire de P à X (en vertu des lois de rétroaction).
- 2) Saillance présente : P est momentanément saillant (en vertu des lois de régulation).
- 3) Motivation existante : la motivation présente pour obtenir P suffit à ce que l'agent se mette en état de produire X.

Notons que la condition 3 met à jour le soubassement fonctionnel de la définition donnée plus haut, selon laquelle la volition est l'effort de réaliser un changement en vertu de cet effort, les conditions 1 et 2 constituant des conditions de possibilité, respectivement générales et occurrentes, de 3. Observons ensuite que les trois conditions énoncées expriment des dispositions mentales et des propriétés occurrentes *qualifiant les relations avec l'environnement*. Elles n'impliquent pas nécessairement de propriété occurrente *extérieure* (la saillance peut être interne), ni non plus de changement corporel (le changement recherché peut être mental, même s'il covarie avec une propriété cérébrale). L'action corporelle est une extension de l'activité mentale en direction des effecteurs, qui crée une forme spécifique de rétroaction visuelle et proprioceptive absente des actions strictement mentales. Dans ce dernier cas, comme par exemple la tentative de se souvenir d'un mot, des rétroactions renseignent aussi l'agent sur ses chances de réussir : ce sont les sentiments épistémiques, comme le sentiment de savoir, ou le sentiment de familiarité.

Ces trois conditions sont nécessaires à toute forme de volition. Elles ne sont pas tirées de la psychologie expérimentale, mais de l'analyse mathématique des systèmes de contrôle adaptatif, et disposent de ce fait d'une parfaite généralité. Notons en particulier qu'elles ne déterminent pas la nature des mécanismes, c'est-à-dire des dispositifs, susceptibles dans chaque espèce organique de satisfaire les contraintes. Elles s'appliquent à un agent quelconque, animal, robot, dès qu'il vise un résultat et en lance l'exécution. Pour montrer que ces conditions sont suffisantes, il faudra vérifier qu'elles sont à l'abri de la déviance causale, dans les limites indiquées plus haut. Nous y reviendrons plus loin.

Il est maintenant possible de répondre succinctement aux deux objections traditionnelles contre le volitionnisme. La définition proposée ne comporte aucune référence à l'action, comme le font certaines théories de la volition qui la *définissent* comme une action mentale, parce que le concept de contrôle est plus général que celui d'action. La condition 3 n'implique pas non plus de circularité : elle n'est dans notre définition qu'une des conditions de l'action, et non l'action elle-même. « La mise en état » n'est pas une action mentale, mais une opération⁶. La définition n'implique pas non plus de régression à l'infini, parce que le contenu d'une volition porte sur l'exécution de X en vue de l'effet P,

6. Sur cette distinction, voir Proust (2001).

et non sur la mise en état d'exécuter X⁷. Le cerveau de l'agent exploite les régularités dont il dispose pour placer celui-ci dans l'état d'exécuter un changement. Cette mise en activité n'est pas décidée par un vouloir antérieur, il constitue le vouloir. Former la volition d'ouvrir la fenêtre n'implique donc pas, en abyme, le recours à une volition pour former cette volition.

Ces précisions sur la théorie volitionniste étant apportées, voyons comment elle peut répondre aux trois difficultés classiques des théories causales. Le problème de Malebranche, on l'a vu, trouve une réponse dans la structure téléologique de la volition. On sait que la structure associant la reproduction à la sélection par les conséquences dispense d'identifier explicitement les raisons qui ont causé la sélection. La loi de l'effet, étant téléologique, ne passe pas non plus par la mémorisation des causes de la sélection. Elle suppose simplement, dans sa version généralisée, que l'agent accède à une portion de l'espace de régulation (qu'il sache comment atteindre une cible quelconque), qu'il ait appris à discerner des saillances, et soit suffisamment motivé par elles.

Dans la conception volitionniste, le problème de la déviance causale reçoit un traitement original. En effet, même si un démon malin intervient dans la commande (soit pour l'initier, soit pour la fausser), l'agent conserve la capacité de se réappropriier ou de rejeter la commande du fait de la structure dynamique de la séquence commande/comparaison intérieure/révision. De deux choses l'une, donc. Soit la volition de l'agent « reprend la main » en quelque sorte, comme le font les agents qui s'adaptent à un changement imprévu du monde, en révisant ou inhibant la commande « manipulée ». Soit la volition reste également manipulée dans le suivi de l'action : l'agent est incapable de réviser et d'inhiber la dérégulation en cours. On doit alors refuser de considérer qu'il s'agit d'une volition, puisque la condition 3 ci-dessus n'est pas satisfaite. Cette fois, le refus n'est pas lié à une convention, mais à la structure réflexive du suivi de la volition.

Enfin, le problème de Wittgenstein trouve une réponse, au demeurant très peu wittgensteinienne, dans une distinction fonctionnelle que la définition de la volition met en évidence : l'action est la volition A d'atteindre le résultat R en vertu de cette volition. C'est, en d'autres termes, la mise en état réflexive occurrente de produire un changement conduisant à R, selon une trajectoire contenue dans l'espace de régulation, avec une motivation occurrente d'atteindre R. Si R fait son apparition, comme l'accélération cardiaque dans l'exemple du yogi, alors qu'il n'existe pas de trajectoire possible dans l'espace de régulation occurrent pour atteindre R, ou si R n'est pas motivant, on ne pourra attribuer de volition d'obtenir R à un agent putatif. En revanche,

7. Il est intéressant de voir que cette dernière formule mobilise un théorème portant sur les conditions générales de régulation dans les systèmes mécaniques. À supposer en effet que la volition V1 suffise à atteindre l'objectif P, il ne peut avoir de volition V2 distincte de V1 qui soit sélectionnée pour atteindre l'effet recherché P, quand V2 n'y parviendrait qu'en activant V1. Car une règle de moindre action gouverne les stratégies de régulation, en vertu d'un théorème dérivant la moindre action de l'existence de symétries (Granger, 1979).

si les trois conditions sont remplies, mais que, pour des raisons étrangères à ces conditions, le résultat R n'apparaît pas, l'action aura effectivement eu lieu, elle sera simplement non réussie.

La définition proposée admet que les volitions peuvent être déclenchées de manière inconsciente. Quel est alors le rôle de la conscience dans l'agir ? L'ouvrage examine trois types de théories. Margaret Anscombe l'identifie à la connaissance non observationnelle de ses intentions, laquelle permet de connaître immédiatement le contenu de ses pensées (conatives ou épistémiques). Le problème est que cette théorie n'explique pas la manière dont s'effectue la prise de conscience, ni par conséquent les illusions ou les erreurs qui peuvent y intervenir. Comment acquérir une connaissance sinon en accédant à une source d'information ? Si la source n'est pas observationnelle, quelle est-elle ? La deuxième théorie est la conception inférentielle défendue entre autres par Dan Wegner, selon laquelle le sentiment d'agir dériverait de l'analyse rétrospective par le sujet d'un certain nombre d'indices prélevés pendant l'exécution de l'action. Cette conception a été expérimentalement mise à l'épreuve et rejetée par Pat Haggard et ses collègues. Haggard montre que la conscience de l'action est sensible à des paramètres temporels qui structurent l'exécution et la perception des effets. D'où la troisième théorie : la conscience de l'action est essentiellement perceptive. On perçoit sa volition d'atteindre un certain but dans le retour perceptif des boucles de contrôle associées. La conscience de l'effort, autre dimension de la conscience d'agir, passe aussi par des marqueurs perceptifs, construits quant à eux par des lois de rétro-action proprioceptive.

Le chapitre six s'intéresse à la question classique du libre-arbitre : la volonté peut-elle être libre ? On y pose la question dont dépend la position même du problème : la liberté est-elle ou non compatible avec le déterminisme des causes qui déterminent la volonté ? Les incompatibilistes répondent par la négative, soit pour rejeter la possibilité conceptuelle de la liberté, soit pour rejeter l'impact des causes sur la volonté. Les compatibilistes en revanche soutiennent qu'on peut trouver un sens à la liberté, que le monde soit ou non déterminé. Cette position a le mérite, du point de vue naturaliste, de rejeter le principe des possibilités alternatives : être libre ne veut pas dire que l'on aurait pu vouloir faire une autre action que celle que l'on a voulu faire. Pour Harry Frankfurt, une volition est libre s'il y a conformité entre elle et une volition de second ordre, la seconde constituant la première comme la volonté de l'agent. Cette théorie séduisante soulève diverses objections, sur la nature des volitions de second ordre, sur le caractère momentané de la liberté du vouloir, et sur la cohérence entre le compatibilisme et la théorie proposée. Frankfurt échappe au risque de la régression à l'infini des niveaux de volonté en trouvant une dimension transversale, qui est la dimension affective, dans laquelle le sujet peut s'unifier. Mais cette solution rend difficile l'adhésion au compatibilisme. La deuxième objection qu'on peut opposer à Frankfurt met en évidence la vulnérabilité de toutes les solutions compatibilistes au subjectivisme : le sujet peut se croire libre sans l'être. Ces observations conduisent à donner du poids à

la troisième objection. Il ne semble pas que le compatibilisme soit une position cohérente si on donne au concept de liberté une dimension non subjectiviste. Ce chapitre évoque en conclusion la possibilité de développer une approche strictement comparative de la liberté de vouloir, qui soit cohérente avec l'incompatibilisme. On dira d'un comportement qu'il est plus libre qu'un autre s'il fait appel à un système de régulation plus flexible, sans occulter le fait que toute décision de régulation est elle-même causalement déterminée.

Le chapitre final se tourne vers une question très familière aux lecteurs de Frankfurt : peut-on vouloir être une personne ? On montre d'abord l'impasse dans laquelle se trouvent les théories qui tentent de construire la personne sur la base de la seule conscience instantanée d'agir, de percevoir, ou d'avoir un corps. Puis on examine les théories mémorielles simples, qui identifient la personne à ce que la personne elle-même se souvient d'elle-même. Ces théories tombent généralement sous le coup de la circularité, et de l'argument leibnizien dit de la « duplication » : elles autorisent deux personnes numériquement distinctes à être la même personne si elles partagent leurs souvenirs. Dans la théorie mémorielle révisée, finalement retenue, l'identité personnelle se fonde non sur le seul recouvrement des souvenirs, mais sur l'exercice de la capacité d'auto-affection. La représentation de soi comme personne identique au fil du temps est la représentation motivante qui émerge du résultat de cet exercice de construction par sélection et révision des valeurs qui déterminent les décisions métacognitives, et ne lui préexiste pas. Mais, une fois en place, elle organise et motive de nouvelles formes de contrôle individuel et social, et de ce fait acquiert un rôle causal.