



ÉCONOMIE

Collection dirigée par Laurent Vidal

DROIT

Causalité, responsabilité et contribution à la dette

Sous la direction de
Samuel Ferey et Florence G'Sell



bruylant





CHAPITRE 3 :

INTENTION, CAUSE, ET RESPONSABILITÉ :

MENS REA ET EFFET KNOBE

Sacha Bourgeois Gironde et Markus Kneer

162. Le droit de la responsabilité civile comme pénale laisse une part importante à l'analyse de l'intentionnalité des comportements qui ont causé les dommages. L'intention permet ainsi de distinguer différents degrés de fautes en droit de la responsabilité civile (fautes intentionnelles/fautes non-intentionnelles) et est également déterminante en droit pénal pour caractériser l'infraction. De même, l'intention joue un rôle dans la répartition finale de la charge de la dette à réparation entre plusieurs coauteurs dès lors que cette charge est partagée au *pro rata* des degrés de faute.

163. De ce point de vue, le dialogue entre le droit et la philosophie de l'action peut s'avérer extrêmement fructueux et s'interroger sur la nature de l'intention d'un point de vue philosophique peut venir utilement compléter et éclairer la théorie comme la pratique du droit. Or, depuis une dizaine d'années, un changement de paradigme se dessine au sein de la théorie de l'action. La conception de l'action intentionnelle, qui prédominait jusqu'alors, a subi de profondes critiques.

164. Plus précisément, un point crucial du consensus en philosophie analytique porte sur l'indépendance conceptuelle de la notion d'intentionnalité à l'égard de considérations de nature morale. Cela signifie que le fait qu'on puisse juger une action comme ayant été produite de manière intentionnelle ou pas est complètement indépendant de ses qualités morales et donc de ses conséquences (349). Le fait de pouvoir

(349) Cette conception standard a été défendue, *inter alia*, par R. J. BUTLER, « Report on Analysis "Problem" no. 16 », *Analysis*, 1978/38, pp. 113-114 ; L. KATZ, *Bad Acts and Guilty Minds: Conundrums of the Criminal Law*, Chicago, University of Chicago Press, 1987 ; A. R. MELE, *Springs of Action: Understanding Intentional Behavior*, Oxford, Oxford University Press, 1992 ; A. R. MELE et P. K. MOSER, « Intentional action », *Nous*, 1994/28, pp. 39-68 ; A. R. MELE, et S. SVERDLIK, « Intention, Intentional Action, and Moral Responsibility », *Philos. Stud.*, 1996/82, pp. 265-287 ; B. F. MALLE et J. KNOBE, « The Folk Concept of Intentionality », *J. Exp. Soc. Psychol.*, 1997/33, pp. 101-121. Des désaccords se sont très tôt déclarés et l'idée que les attributions d'intentionnalité *sont* de fait, ou





déterminer si un agent effectue une action intentionnellement joue alors un rôle décisif dans l'appréciation de son éventuelle culpabilité et de l'étendue de cette culpabilité. Ce consensus se doublait d'une position méthodologique consistant à affirmer que le raisonnement analytique, la philosophie « en fauteuil », était le mode de travail le plus approprié pour justifier une théorie de l'action intentionnelle.

165. Or, à la suite de plusieurs travaux de Joshua Knobe, cette position consensuelle a été critiquée. Pour Joshua Knobe, il est possible de renverser l'inférence et de considérer qu'intention et conséquences ne sont pas indépendantes. Dit autrement, on infère souvent l'intentionnalité de la culpabilité. Cette position s'appuie parallèlement sur un renouvellement méthodologique qui soumet les concepts de portée philosophique à une approche empirique, la philosophie expérimentale (350). Ainsi, c'est en interrogeant des individus ordinaires sur différents types de *scenarii* construits que Joshua Knobe identifie cet effet, appelé depuis « effet Knobe » ou « *side-effect effect* ». Sa conclusion est alors de montrer que le concept d'action intentionnelle, tel qu'il est formulé et utilisé en philosophie, et le concept populaire, dont la philosophie est pourtant censée rendre compte, sont assez éloignés l'un de l'autre et ce, alors même que la définition de la philosophie analytique consiste en l'explication des concepts véhiculés par le langage ordinaire.

166. À la suite de la mise en évidence de l'effet Knobe, un nombre croissant d'articles théoriques et expérimentaux a vu le jour sur les attributions d'intentionnalité et d'autres états mentaux. Notre intention présente n'est pas de proposer une nouvelle théorie du mécanisme sous-jacent à l'effet observé, mais plutôt d'explorer l'étendue et la portée de ce phénomène dans un contexte où, selon nous, il revêt une importance cruciale : le droit et la pratique juridique. Il y a une forte tension entre le concept d'intentionnalité, tel qu'il est usité dans la théorie de l'action, en philosophie du droit et dans la théorie juridique, et le concept naïf véhiculé par le langage et la pensée ordinaires. Étant donné le caractère décisif d'une attribution d'intentionnalité dans le processus d'établissement de la conviction qu'un accusé a agi avec une *mens rea* (une

devraient être, sensibles à des considérations morales a été défendue notamment par G. HARMAN, « Practical Reasoning », *Rev. Metaphys.*, 1976/29, pp. 431-463 ; E. J. LOWE, « Neither Intentional nor Unintentional », *Analysis*, 1978/38, pp. 117-118 ; A. DUFF, « Intention, Responsibility and Double Effect », *Philos. Q.*, 1982/32, pp. 1-16 ; M. BRATMAN, *Intention, Plans, and Practical Reason*, coll. David Hume Series, Cambridge, Harvard University Press, 1987.

(350) Le développement de la philosophie expérimentale a entraîné dans son sillage des débats de fond sur les potentialités et les limites de la philosophie « en fauteuil », sur l'utilité des résultats des questionnaires de la psychologie naïve et, plus généralement, sur le rôle des intuitions en philosophie.





intention criminelle), il paraît impératif de déterminer lequel des deux concepts d'intentionnalité, celui qui relève de la philosophie ou de la psychologie naïve, les professionnels du droit, et en particulier les juges, ont tendance à employer. En clair, le but de cette contribution est de présenter et de discuter des données expérimentales sur la question de l'usage par les professionnels du droit du concept d'intentionnalité et de déterminer si cet usage est sensible aux considérations morales (comme chez les sujets naïfs) ou non affecté par celles-ci (comme pour les philosophes et les théoriciens du droit).

167. Le reste de cet article procède de la manière suivante. Dans la première section nous présentons de manière précise l'effet Knobe et en discutons les principales explications. Leur pertinence relativement à la pratique juridique est examinée plus précisément dans la deuxième section. Dans la troisième section nous présentons des données expérimentales sur l'effet Knobe acquises sur un échantillon de praticiens du droit (avocates et universitaires). La quatrième section de cet article résume rapidement les résultats d'une étude similaire par les auteurs, sondant, cette fois, les intuitions de juges français. La cinquième et dernière section résume et souligne les leçons, éventuellement préoccupantes, qui peuvent être tirées de ces résultats expérimentaux.

SECTION 1 – L'EFFET KNOBE ET SON EXPLICATION

L'approche et la conception traditionnelles ont été battues en brèche par les recherches empiriques de Joshua Knobe qui ont porté sur ce qu'il est convenu d'appeler « l'effet-effet secondaire », pour lequel nous nous permettons de conserver la dénomination anglaise plus naturelle de « *side-effect effect* ».

1. EXPOSÉ DE L'EFFET KNOBE

168. Joshua Knobe a mis au point un scénario hypothétique dans lequel une action principale intentionnelle (par exemple mettre en place une stratégie commerciale dans le but d'augmenter ses profits) donne lieu à des effets secondaires opposés (351). Dans l'une des deux versions du scénario, cette nouvelle stratégie commerciale génère des nuisances environnementales, dans l'autre, elle est au contraire bénéfique

(351) J. KNOBE, « Intentional Action and Side Effects in Ordinary Language », *Analysis*, 2003/63, pp. 190-194.



à l'environnement. La question est alors de savoir ce que va être l'intuition du sujet naïf eu égard au caractère intentionnel de l'effet secondaire de l'action engagée en fonction de la valence de cette action. La moitié des participants de l'expérience de Joshua Knobe étaient ainsi soumis à la vignette suivante :

« Le vice-président d'une entreprise va voir le président et dit : 'Nous avons mis au point un nouveau programme. Celui-ci nous permettra d'augmenter considérablement nos profits mais il aura aussi pour effet de nuire à l'environnement'. Le président répond : 'Je me fiche complètement de nuire à l'environnement. Tout ce qui m'intéresse, c'est de faire le plus de profits possibles. Démarrons ce nouveau programme'. Le nouveau programme est démarré. Comme prévu, l'environnement en pâtit. » (352)

169. Les participants étaient interrogés sur le fait de savoir si le président de la compagnie a intentionnellement ou pas nuï à l'environnement et étaient soumis à un choix binaire contraint : oui ou non. Dans cette version où l'effet secondaire correspond à une nuisance, 82 % des participants ont répondu que l'effet secondaire était produit de manière intentionnelle.

170. L'autre moitié des participants devaient considérer un scénario quasi-identique, à ceci près que l'effet secondaire ne consistait plus cette fois en une nuisance environnementale mais à un effet bénéfique (les différences entre les deux versions sont soulignées ici) :

« Le vice-président d'une entreprise va voir le président et dit : 'Nous avons mis au point un nouveau programme. Celui-ci nous permettra d'augmenter considérablement nos profits, et il aura aussi pour effet d'aider l'environnement'. Le président répond : « Je me fiche complètement d'aider l'environnement. Tout ce qui m'intéresse, c'est de faire le plus de profits possibles. Démarrons ce nouveau programme'. Le nouveau programme est démarré. Comme prévu, l'environnement en bénéficie » (353).

171. Dans cette condition bénéfique, seuls 23 % des participants ont répondu que le président avait contribué à la qualité de l'environnement de manière intentionnelle. Le sujet naïf, ainsi donc, tend à considérer que les effets secondaires négatifs dont le sujet de l'action principale est conscient sont intentionnels, tandis que des effets secondaires positifs

(352) *Ibid.*, p. 191.

(353) *Ibid.*, p. 191.



sont jugés comme des à-côtés non intentionnels de l'action principale. Cet effet résiste remarquablement à toutes sortes de réplifications et il s'avère robuste à travers les cultures et l'âge des sujets (354). D'autres études ont également montré que la différence dans la valence des effets secondaires suscitait des jugements asymétriques similaires quand il s'agissait de demander aux participants si le résultat était *désiré*, si l'agent avait *décidé* de produire ce résultat, et s'il *croyait* ou *savait* qu'il produirait ce résultat (355).

172. La différence entre les deux versions du scénario de Joshua Knobe ne tient *a priori* qu'à la valence morale opposée, d'un cas à l'autre, de l'effet secondaire provoqué par l'action principale ordonnée par le président de la compagnie. Dans la vignette négative, cette action principale génère un effet secondaire moralement indésirable (l'environnement est pollué), dans la vignette positive, cette même action donne lieu à un effet secondaire moralement désirable (la qualité de l'environnement est améliorée). Pourquoi une telle différence dans la valence morale d'un effet secondaire suscite des attributions d'intentionnalité radicalement asymétriques ?

(354) L'effet a été largement répliqué. Voy. J. KNOBE, « Intentional action and side effects in ordinary language », *op. cit.*, p. 191 ; J. KNOBE, « Intentional Action in Folk Psychology: An Experimental Investigation », *Philos. Psychol.*, 2003/16, pp. 309-324 ; J. KNOBE, « Folk Psychology and Folk Morality: Response to Critics », *Journal of Theoretical and Philosophical Psychology*, 2004/24, pp. 270-279 ; J. KNOBE, « Intention, Intentional Action and Moral Considerations », *Analysis*, 2004/64, pp. 181-187 ; J. KNOBE et G.S. MENDLOW, « The Good, the Bad and the Blameworthy: Understanding the Role of Evaluative Reasoning in Folk Psychology », *Journal of Theoretical and Philosophical Psychology*, 2004/24, pp. 252-258 ; T. NADELHOFFER, « Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow », *Journal of Theoretical and Philosophical Psychology*, 2004/24, pp. 259-269 ; T. NADELHOFFER, « Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality », *Philos. Explor.*, 2006/9, pp. 203-219 ; H. J. MCCANN, « Intentional Action and Intending: Recent Empirical Studies », *Philos. Psychol.*, 2005/18, pp. 737-748 ; A. R. MELE, et F. CUSHMAN, « Intentional Action, Folk Judgments, and Stories: Sorting Things Out », *Midwest Studies in Philosophy*, 2007/31, pp. 184-201. Il résiste aux différences culturelles (J. KNOBE et A. BURRA, A., « The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study », *J. Cogn. Cult.*, 2006/6, pp. 113-132) et aux différences d'âge (A. M. LESLIE, J. KNOBE et A. COHEN, « Acting Intentionally and the Side-Effect Effect Theory of Mind and Moral Judgment », *Psychol. Sci.*, 2006/17, pp. 421-427). Pour une présentation générale, voy. A. FELTZ, « The Knobe effect: A brief Overview », *J. Mind Behav.*, 2007/28, pp. 265-277.

(355) En ce qui concerne les désirs, voy. D. TANNENBAUM, P. H., DITTO et D. A. PIZARRO, « Different Moral Values Produce Different Judgments of Intentional Action », *Unpublished manuscript, University of California-Irvine*, 2007 ; pour les décisions, plusieurs auteurs mettent en évidence un *side-effect effect* épistémique pour ce qui est de la croyance. Voy. D. PETTIT et J. KNOBE, « The Pervasive Impact of Moral Judgment », *Mind Lang.*, 2009/24, pp. 586-604 ; J. R. BEEBE, « A Knobe Effect for Belief Ascriptions », *Rev. Philos. Psychol.*, 2013/4, pp. 235-258 et M. KNEER, « Beyond Bivalence: A Graded Model of the Knobe Effect ? », unpublished manuscript, 2015). Pour la connaissance, voy. J. R. BEEBE et W. BUCKWALTER, « The Epistemic Side-Effect Effect », *Mind Lang.*, 2010/25, pp. 474-498 ; J. R. BEEBE et M. JENSEN, « Surprising Connections between Knowledge and Action: The Robustness of the Epistemic Side-Effect Effect », *Philos. Psychol.*, 2012/25, pp. 689-715.



173. Une des premières explications de ce phénomène qui a été proposée évoque le fait que « les intuitions qu’entretiennent les gens à l’égard d’une action intentionnelle tendent à se fixer sur les traits psychologiques qui sont les plus directement pertinents en vue de l’expression de la louange ou du blâme » (356). Toutes les mauvaises actions ne suscitent pas l’expression du blâme et toutes les bonnes actions n’éveillent pas la louange. Les jugements de blâme et de louange dépendent, entre autres choses, du point de savoir si l’agent pouvait *prévoir* le résultat de son action, s’il *essayait* de provoquer ce résultat, s’il avait la *capacité* ou l’*aptitude* de produire ce résultat. Dans cette veine, Joshua Knobe énonce l’argument que « divers traits psychologiques pourront s’avérer pertinents selon que le comportement lui-même est bon ou mauvais » (357). L’hypothèse est ici que cette asymétrie relative à la pertinence des traits est ce qui influence les attributions d’intentionnalité de la même manière qu’elle affecte les jugements de louange et de blâme.

174. Sous cette approche, la procédure d’attribution de l’intentionnalité peut être subdivisée en deux étapes. Dans la première étape, on évalue si l’action est bonne ou mauvaise. Cette évaluation détermine quels sont les traits du comportement qui sont pertinents en vue de l’attribution de l’intentionnalité, car l’impact de ces traits est variable selon la valence morale de l’action. Lors de la seconde étape, l’évaluation consiste à savoir si le comportement exhibe en fait les traits jugés pertinents pour l’attribution de l’intentionnalité, et un verdict sur ce caractère intentionnel est alors posé. Par exemple, dans le cas d’une action négative telle qu’une atteinte à l’environnement, l’anticipation, de la part de l’agent, semble une condition suffisante en vue de l’attribution du caractère intentionnel, de même qu’elle paraît suffisante pour émettre le blâme. Mais comme le montre ce premier scénario de Joshua Knobe, ce sont d’autres conditions qui s’appliquent apparemment pour les conséquences positives telles que l’amélioration de l’environnement. Dans ce cas, l’anticipation, en elle-même, ne paraît pas suffisante pour émettre une louange et attribuer un caractère intentionnel à cette conséquence de l’action du chef d’entreprise – un désir de favoriser cette conséquence ou une tentative de la faire aboutir paraissent devoir être requis à cette fin.

175. Cette première explication fournie par Joshua Knobe n’a pas résisté à un certain nombre de nouvelles données expérimentales. Dans

(356) J. KNOBE, « The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology », *Philos. Stud.*, 2006/130, pp. 203-231, p. 225.

(357) *Ibid.*, p. 225.



un scénario alternatif (dit des « ventes du *New Jersey* »), un dirigeant d'entreprise adopte de façon délibérée une nouvelle stratégie commerciale, qui a pour effet d'augmenter les profits de l'entreprise mais également d'augmenter/diminuer les ventes dans l'État du *New Jersey*. Les données reproduisent une asymétrie significative dans les attributions d'intentionnalité dans les versions positives et négatives : la diminution des ventes était jugée intentionnelle, mais pas leur augmentation (358). L'effet secondaire négatif n'est pas moralement mauvais dans un sens obvie et les participants interrogés sur ce point ne le jugeaient pas comme tel. Ils ne pensaient pas non plus que le dirigeant, que ce soit dans la version positive ou la version négative, ne méritait d'être respectivement loué ou blâmé. Ces résultats suggèrent ainsi, en premier lieu, que les considérations morales – essentielles à la première étape décrite par Joshua Knobe dans la procédure d'attribution de l'intentionnalité – ne revêtent pas une importance fondamentale dans l'explication de l'effet. En second lieu, ces résultats remettent également directement en cause l'idée que « les intuitions qu'entretiennent les gens à l'égard d'une action intentionnelle tendent à se fixer sur les traits psychologiques qui sont les plus directement pertinents en vue de l'expression de la louange ou du blâme » (359). Car ce second scénario n'implique tout simplement pas de jugements de blâme ou de louange et ne manifeste aucun trait qui leur soit pertinent (360).

176. Dans des articles récents, Joshua Knobe propose une nouvelle théorie qui prend en compte une quantité impressionnante de données sur le *side-effect effect* (361). Des scénarios mettant en jeu des effets

(358) Voy. M. PHELAN et H. SARKISSIAN, « The Folk Strike Back; or, Why you Didn't Do It Intentionally, Though it Was Bad and You Knew It », *Philos. Stud.*, 2008/138, pp. 291-298, le scénario a été conçu par Knobe and Mendlow (J. KNOBE et G.S. MENDLOW, « The Good, the Bad and the Blameworthy : Understanding the Role of Evaluative Reasoning in Folk Psychology », *op. cit.*).

(359) J. KNOBE, « The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology », *op. cit.*, p. 225.

(360) Selon l'explication proposée, si le résultat est mauvais, l'anticipation de ce résultat est *suffisante* pour une attribution d'intentionnalité. Toutefois, cette explication s'avère fautive (voy. par exemple le scénario de la mare dans A. R. MELE, et F. CUSHMAN, « Intentional Action, Folk Judgments, and Stories: Sorting Things Out », *op. cit.*). Dans certains scénarios, dans lesquels l'agent anticipe l'effet secondaire négatif mais exprime des regrets à son propos, cette conséquence de son action n'est pas jugée intentionnelle. Pour des expériences de ce type, voy. S. SVERDLIK, « Intentionality and Moral Judgments in Commonsense Thought About Action », *Journal of Theoretical and Philosophical Psychology*, 2004/24, pp. 224-236. L'expérience du planificateur urbain de M. T. Phelan and Sarkissian (M. PHELAN et H. SARKISSIAN, « The folk strike back; or, why you didn't do it intentionally, though it was bad and you knew it », *op. cit.*) et le scénario du levier de Lanteri (A. LANTERI, « Judgements of Intentionality and Moral Worth: Experimental Challenges to Hindriks », *Philos. Q.*, 2009/59, pp. 713-720). À la suite de ces expériences, Knobe a modifié sa position.

(361) J. KNOBE, « Person as Scientist, Person as Moralist », *Behav. Brain Sci.*, 2010/33, pp. 315-329 ; D. PETTIT et J. KNOBE, « The Pervasive Impact of Moral Judgment », *op. cit.*



secondaires de valence opposée ne donnent pas seulement lieu à des attributions asymétriques d'intentionnalité, mais affectent également les intuitions relatives au fait de juger si l'agent a désiré, soutenu, favorisé cette conséquence, ou encore s'il l'a décidé, s'il y croyait, s'il savait que cela se passerait de cette manière :

« À la lumière de ces résultats, nous sommes enclins à penser que l'influence du jugement moral est diffuse, qu'elle joue un rôle dans l'application de tout concept qui a trait au fait de tenir ou manifester une attitude positive vis-à-vis d'une conséquence. Autrement dit, pour tout concept de cette sorte, nous formons l'hypothèse qu'il y a un processus psychologique qui pousse les gens à attribuer plus volontiers le concept en question dans le cas d'effets secondaires moralement négatifs et moins volontiers dans le cas d'effets secondaires positifs. » (362)

177. L'hypothèse se généralise à une variété de concepts psychologiques ordinaires, car tous présentent une caractéristique commune : ils sont sensibles à différents critères d'attribution par défaut selon la valence de l'effet secondaire considéré. Ce point peut être illustré en observant le comportement d'adjectifs quantifiables (*gradable adjectives*) tels que froid, grand ou éduqué. Par exemple, les jugements portant sur la température de diverses boissons impliquent différents critères. Un café à 20° C est considéré comme froid, parce qu'il est plus froid que sa température par défaut lorsque l'on jouit de sa consommation. De même une bière à 20° C sera jugée tiède, car elle est alors au-dessus de sa température par défaut. Selon la boisson, le critère par défaut de ce qui peut être jugé chaud ou froid varie.

178. Une logique similaire gouverne l'attribution des expressions du type « intentionnel », « désiré » ou « est en faveur de ». Les attentes par défaut concernant les propriétés contextuelles et psychologiques qui gouvernent, par exemple, les attributions d'intentionnalité varient ainsi selon que les actions concernées donnent lieu à des effets secondaires positifs ou négatifs. L'asymétrie dans l'attribution d'intentionnalité dans le premier scénario lié à la pollution ou à l'amélioration de l'environnement s'explique ainsi *exclusivement* par un glissement dans *les attentes standards* : lorsque l'effet secondaire est négatif, l'anticipation consciente de cet effet par l'agent motive l'attribution d'intentionnalité, quand cet effet est positif, le critère est plus strict et une attitude

(362) J. KNOBE, « Folk judgments of causation », *Studies in History and Philosophy of Science Part A*, 2009/40, pp. 238-242, spéc. p. 238.



plus proactive de la part de l'agent est requise pour dire que cette action est intentionnelle. Cette explication, affirme Joshua Knobe,

« postule un rôle pour une sorte de jugement moral complètement à part – un jugement qui peut être formulé y compris en l'absence de toute information concernant cet agent particulier ou son comportement. L'idée est donc que, avant même que les gens aient pu commencer à considérer ce qui s'est réellement passé dans la situation, ils portent leur attention sur l'action de nuire à l'environnement et forme un jugement sur le type d'attitude qu'on peut attendre d'un agent vis-à-vis d'une telle conséquence. Ce jugement leur sert de critère qu'ils peuvent utiliser pour rendre compte du comportement qu'ils observent. » (363)

2. UNE TYPOLOGIE DES EXPLICATIONS

179. Prenons un peu de recul afin d'isoler quelques aspects généraux de ces possibles explications. Elles peuvent être conceptualisées selon trois dimensions principales. La première dimension tient au fait de savoir si le *side-effect effect* révèle un aspect essentiel, mais jusqu'ici méconnu, de la psychologie humaine, ou s'il doit être bien plutôt considéré comme un biais affectant la cognition ordinaire. Nous suivrons la littérature sur ce point en désignant la première hypothèse d'explication en termes de « compétence » et la seconde d'explication en termes de « biais ».

180. La seconde dimension concerne la question de savoir s'il y a bien un concept *unique* d'intentionnalité, ou si les résultats de Joshua Knobe mettent en fait en lumière un concept *multiple* d'intentionnalité. L'un ou l'autre choix sur cette dimension n'exclut de fait aucune position sur la dimension précédente. Les données peuvent tout à fait être interprétées en faveur d'un concept unique appliqué de manière compétente ou non, ou alors d'un concept multiple appliqué de manière compétente ou non. Une position intermédiaire peut être également envisagée : si l'expression « intentionnellement » est employé de manière diverse et variée, *certain*s de ses emplois peuvent être jugés légitimes et *d'autres*, en dépit de leur caractère répandu, défectueux et incohérents.

181. Une dernière dimension différencie les théories entre elles selon le type d'état psychologique invoqué dans l'explication du *side-effect*

(363) J. KNOBE, « Person as Scientist, Person as Moralist », *op. cit.*, p. 328.





effect. La question centrale consiste ici à savoir si l'asymétrie exprimée par les sujets met en œuvre des croyances descriptives ou des attitudes normatives vis-à-vis de l'agent du scénario. Dans le reste de cette section, nous allons rapidement passer en revue la littérature pour ce qui concerne la seconde et la troisième dimension d'analyse, et nous reviendrons ensuite sur la différence entre compétence et biais qui est en fait la plus essentielle dans l'établissement de notre argument d'ensemble.

1 – *Concept unifié ou multiple*

182. Une question centrale est donc de savoir si nous avons affaire à un concept simple et unifié d'intentionnalité, ou si les participants aux expériences font usage de multiple concepts (ou conceptions) d'intentionnalité. Dans une expérience menée par Shaun Nichols et Joseph Ulatowski, les participants devaient considérer conjointement la version positive et la version négative du scénario environnemental (364). Ils devaient répondre par oui ou par non à la question de savoir si le président avait intentionnellement nui ou contribué à l'environnement, et expliquer les motifs de leur réponse. Les données de cette expérience indiquent qu'un tiers des participants répondaient affirmativement aux deux versions, un autre tiers négativement aux deux versions, et un dernier tiers répondait affirmativement à la version avec pollution de l'environnement et négativement à la version avec amélioration de l'environnement. Les sujets qui répondaient affirmativement aux deux versions du scénario justifiaient leur réponse de manière cohérente en faisant référence au fait que le président agissait en pleine conscience des effets secondaires de son action. Les sujets qui répondaient négativement aux deux versions, différemment, soulignaient le fait que le président ne désirait pas ces effets secondaires et que cela suffisait pour ne pas lui attribuer une intentionnalité à leur égard. La prise en compte de cette diversité interprétative pourrait favoriser, d'une part, une conception de l'intentionnalité, légitime, basée sur le concept de connaissance ou de conscience, et, d'autre part, une conception, également légitime, basée sur le concept de motivation ou de désir.

183. Dans leur discussion critique de cet article Fiery Cushman et Alfred Mele proposent un ancrage conceptuel supplémentaire (365). À partir de nouvelles données, ces auteurs suggèrent que le concept

(364) S. NICHOLS et J. ULATOWSKI, « Intuitions and Individual Differences: The Knobe Effect Revisited », *Mind Lang.*, 2007/22, pp. 346-365.

(365) F. CUSHMAN et A. R. MELE, « Intentional Action », in *Experimental Philosophy* (J. KNOBE et S. NICHOLS eds), Oxford, Oxford University Press, 2008, pp. 171-188.





d'intentionnalité est en réalité tripartite et qu'il faut en sus inclure le « [concept] selon lequel la croyance est une condition suffisante pour l'intentionnalité de l'action » (366) et un autre concept pour lequel « la croyance est une condition suffisante pour l'intentionnalité de l'action dans les seuls cas d'actions moralement négatives » (367). De son côté, Joshua Knobe, à l'instar de la majorité des approches en vigueur, défend l'unicité conceptuelle de l'intentionnalité. La sensibilité contextuelle du terme « intentionnel » chez Joshua Knobe à l'origine d'attributions asymétriques selon des contextes de différentes valences, ne constitue pas davantage en soi un argument en faveur de son ambiguïté conceptuelle, que la sensibilité d'adjectifs quantifiables comme « froid » lorsqu'il est appliqué à diverses boissons ne montrent que « froid » est un concept ambigu.

2 – *Considérations descriptives vs. normatives*

184. Les théories du *side-effect effect* portent également sur le type d'états psychologiques des participants au moment de leurs jugements et qui permettrait d'en expliquer l'asymétrie. Certaines explications, ainsi, se concentrent prioritairement sur les croyances de nature descriptive que nous entretenons au sujet de l'agent lorsque nous émettons notre jugement d'attribution d'intentionnalité ; d'autres, différemment, se focalisent sur les attitudes normatives que nous formons à l'égard de l'agent du scénario. Dans la première approche du problème que Joshua Knobe avait développée, par exemple, son explication est clairement normative, dans la mesure où la louange et le blâme y jouent un rôle déterminant. Il en va de même dans sa seconde approche, selon laquelle une évaluation morale du comportement de l'agent pose les conditions implicites par défaut relativement auxquelles s'effectuent les attributions d'intentionnalité. F. Hindriks, de son côté, avec son « hypothèse de raison normative », propose que, lorsqu'il se trouve face à l'éventualité d'un effet secondaire négatif, un agent a une bonne raison normative de l'éviter, et de suspendre l'action principale susceptible de la causer (368). S'il omet de le faire, l'effet secondaire peut être jugé intentionnel. Mais pour ce qui est des effets secondaires positifs, aucune raison

(366) *Ibid.*

(367) *Ibid.* et A. LANTERI, « Three-and-a-half Folk Concepts of Intentional Action », *Philos. Stud.*, 2012/158, pp. 17-30, pour un quatrième concept d'intentionnalité.

(368) F. HINDRIKS, « Intentional Action and the Praise-Blame Asymmetry », *Philos. Q.*, 2008/58, pp. 630-641.



normative de la sorte n'entre en jeu, et, par conséquent, l'effet secondaire n'est pas jugé intentionnel. (369)

185. Edouard Machery soutient pour sa part une alternative descriptiviste aux approches normatives dont nous venons de parler (370). Dans la condition négative du scénario environnemental, l'agent fait face à un arbitrage coût-bénéfice entre l'augmentation des profits et les dommages infligés à l'environnement. Un tel arbitrage est absent de la version positive. Le fait d'encourir des coûts anticipés, selon Edouard Machery, peut être considéré comme un acte intentionnel, ce qui ne serait pas le cas en cas de bénéfices provenant d'effets secondaires, même anticipés. Dans la mesure où les arbitrages coûts-bénéfices n'ont pas de dimension morale intrinsèque, cette hypothèse d'arbitrage constitue, dans la typologie des approches du *side-effect effect*, qui tombe dans la catégorie des explications descriptivistes et basées sur la compétence des agents (371).

3 – *Compétence vs. Biais*

186. Le raisonnement, le jugement et le choix humains sont affectés par de nombreux biais. Par exemple, des options identiques induisent des choix différents selon la manière dont elles sont « cadrées » ou présentées (372). Les gens tendent également à surestimer leur contribution positive et à sous-estimer leur contribution négative à des activités collectives (*self-serving bias*). Le biais de conservation consiste à sélectionner les faits qui vont confirmer, et à tendre à ignorer les faits qui vont contredire, une hypothèse préalablement adoptée (373).

(369) Une proposition intéressante, bien qu'un peu plus complexe, et invoquant des normes et des heuristiques doxastiques est faite par M. ALFANO, J. R. BEEBE et B. ROBINSON, « The Centrality of Belief and Reflection in Knobe-Effect Cases », *Monist*, 2012/95, pp. 264-289.

(370) E. MACHERY, « The Folk Concept of Intentional Action: Philosophical and Experimental Issues », *Mind Lang.*, 2008/23, pp. 165-189.

(371) Pour une critique de cette hypothèse d'arbitrage, voy. R. MALLON, « Knobe versus Machery: Testing the Trade-off Hypothesis », *Mind Lang.*, 2008/23, pp. 247-255 et M. PHELAN et H. SARKISSIAN, « Is the 'Trade-off Hypothesis' Worth Trading For ? », *Mind Lang.*, 2009/24, pp. 164-180. Pour une approche descriptiviste, voy. C. S. SRIPADA, « The Deep Self Model and Asymmetries in Folk Judgments about Intentional Action », *Philos. Stud.*, 2010/151, pp. 159-176 et C. S. SRIPADA et S. KONRATH, « Telling More than we Can Know about Intentional Action », *Mind Lang.*, 2011/26, pp. 353-380, discutés de manière critique par F. COVA et H. NAAR, « Testing Sripada's Deep Self Model », *Philos. Psychol.*, 2012/25, pp. 647-659. Pour une approche mixte, voy. R. HOLTON, « Norms and the Knobe Effect », *Analysis*, 2010/70, pp. 1-8.

(372) A. TVERSKY et D. KAHNEMAN, « The Framing of Decisions and the Psychology of Choice », *Science*, 1981/211, pp. 453-458.

(373) Voy. P. C. WASON, « On the Failure to Eliminate Hypotheses in a Conceptual Task », *Q. J. Exp. Psychol. Quarterly*, 1960/12, pp. 129-140 ; A. KORiat, S. LICHTENSTEIN et B. FISCHHOFF, « Reasons for confidence », *J. Exp. Psychol.-Learn. Mem. Cogn.*, 1980/6, pp. 107-118 ; D. N. PERKINS, R. ALLEN et J. HAFNER, « Difficulties in Everyday Reasoning », *Thinking : The Expanding Frontier*

Le point central que nous souhaitons aborder est celui de savoir si l'asymétrie dans les jugements d'attribution d'intentionnalité que nous observons est la manifestation d'une compétence des sujets relativement au concept d'intentionnalité ou au contraire d'une incompétence (auquel cas on dira qu'ils sont victimes d'un biais).

187. Bien que l'hypothèse de la compétence prédomine dans cette discussion, les premières données acquises par Joshua Knobe ont suscité une variété d'explications en termes de biais. Face aux difficultés rencontrées par l'explication en termes de compétence et devant le flux de données, l'explication alternative selon laquelle l'attribution de concepts psychologiques ordinaires est affectée par des biais systématiques a acquis une certaine popularité (374). Fred Adams et Annie Steadman expliquent l'asymétrie comme une erreur de performance linguistique (375). Les expériences de Joshua Knobe mettraient en jeu « la dimension pragmatique du discours autour de l'intentionnalité » (376) mais ne toucheraient en rien au concept d'action intentionnelle proprement dit. Fred Adams et Annie Steadman suggèrent que, contrairement aux apparences, les gens ne pensent pas réellement que le président de la compagnie a intentionnellement nui à l'environnement. Ce qu'ils veulent faire, en revanche, est exprimer leur opinion qu'il doit

(W. MAXWELL dir.), Philadelphia, Franklin Institute Press, 1983, pp. 177-189. Pour une recension, voy. J. KLAYMAN, « Varieties of Confirmation Bias », *Psychol. Learn Motiv.*, 1995/32, pp. 385-418 ; R. S. NICKERSON, « Confirmation Bias: A Ubiquitous Phenomenon in many Guises », *Rev. Gen. Psychol.*, 1998/2, pp. 175-220.

(374) Des explications du *side-effect effect* en termes de biais ont été proposées, entre autres, par A. R. MELE, « Acting Intentionally: Probing Folk Notions », in *Intentions and intentionality: Foundations of social cognition* (B. F. MALLE, L. J. MOSES, D. A. BALDWIN eds), Cambridge, MIT Press, 2001, pp. 27-43 et par B. F. MALLE et S. NELSON, « Judging *Mens Rea*: The Tension between Folk Concepts and Legal Concepts of Intentionality », *Behavioral Sciences and the Law*, 2003/21, pp. 563-580 ; F. ADAMS et A. STEADMAN, « Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding ? », *Analysis*, 2004/64, pp. 173-181 ; F. ADAMS et A. STEADMAN, « Intentional Action and Moral Considerations: still Pragmatic », *Analysis*, 2004/64, pp. 268-276 ; T. NADELHOFFER, « Blame, Badness, and Intentional Action: A Reply to Knobe and Mendlow », *op. cit.* ; T. NADELHOFFER, « On Praise, Side Effects, and Folk Ascriptions of Intentionality », *Journal of Theoretical and Philosophical Psychology*, 2004/24, pp. 196-213 ; T. NADELHOFFER, « Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality », *op. cit.* ; J. NADO, « Effects of Moral Cognition on Judgments of Intentionality », *Br. J. Philos. Sci.*, 2008/59, pp. 709-731. Un biais bien connu en liaison avec des phénomènes voisins à celui qui nous intéresse est la théorie de la « causalité coupable » et du « contrôle coupable » proposée par Alicke. Voy. M. D. ALICKE, « Culpable Causation », *J. Pers. Soc. Psychol.*, 1992/63, pp. 368-378 ; M. D. ALICKE, « Culpable Control and the Psychology of Blame », *Psychol. Bull.*, 2000/126, pp. 556-574 ; M. D. ALICKE, « Blaming Badly », *J. Cogn. Cult.*, 2008/8, pp. 179-186. Pour une discussion stimulante récente sur la raison pour laquelle les explications en termes de biais sont certainement sur la bonne piste, on peut se reporter à H. SAUER et T. BATES, « Chairmen, Cocaine, and Car Crashes: The Knobe Effect as an Attribution Error », *J. Ethics*, 2013/17, pp. 305-330.

(375) F. ADAMS et A. STEADMAN, « Intentional Action and Moral Considerations: still Pragmatic », *op. cit.*

(376) *Ibid.*, p. 174.

être blâmé pour son comportement. En l'absence d'un moyen adéquat disponible pour exprimer cette opinion, ils attribuent une intentionnalité au président dans le but de générer l'implication conversationnelle selon laquelle celui-ci est blâmable. Et dans la mesure où les gens n'ont pas tendance à croire que la conscience de produire un effet secondaire positif est digne de louange morale, ils n'attribuent pas de caractère intentionnel dans la version positive du scénario de sorte à éviter une semblable implication pragmatique (377).

188. Outre un tournant linguistique dans l'interprétation du *side-effect effect*, on peut aussi noter des approches de teneur psychologique. Dans ces approches, le concept d'intentionnalité, rigoureusement compris, n'est pas sensible à des considérations d'ordre moral, et son application ne doit donc pas l'être davantage. Toutefois, des considérations morales interfèrent de fait avec l'attribution naïve de l'intentionnalité induisant, par-là, de nombreuses applications erronées du concept.

189. Une conception bien connue en psychologie morale soutient que le « jugement moral est avant tout une affaire d'émotion et d'intuition affective plutôt que de jugement délibératif » et qu'il évoque en priorité « des capacités de l'esprit à résoudre de nombreux problèmes, y compris de haut niveau, de manière inconsciente et automatique » (378). Pour Thomas Nadelhoffer, ces processus non-rationnels, inconscients et affectifs ne sous-tendent pas seulement les jugements moraux mais aussi nos attributions d'intentionnalité (379) :

(377) Dans la mesure où l'effet-Knobe survient également dans des scénarios où la moralité, au sens strict, ni le blâme et le louange, ne jouent de rôle (voy. l'argument ci-dessus concernant les ventes du New Jersey), cette approche a une portée relativement limitée. Voy. cependant S. GUGLIELMO et B. F. MALLE, « Can Unintended Side Effects be Intentional? Resolving a Controversy over Intentionality and Morality », *Pers. Soc. Psychol. Bull.*, 2010/36, pp. 1635-1647. Pour une explication de l'asymétrie en raison de traits linguistiques qui ne repose pas sur des considérations liées au blâme ou la louange.

(378) J. GREENE et J. HAIDT, « How (and where) Does Moral Judgment Work? », *Trends Cogn. Sci.*, 2002/6, pp. 517-523, p. 517.

(379) Cette théorie affective du jugement moral sous-tend le modèle du contrôle coupable de l'attribution du blâme proposée par Alicke. Ce modèle explique les attributions asymétriques de contrôle et d'implication causale en fonction de la valence morale, un phénomène extrêmement semblable à l'effet Knobe : « Quand le mode d'attribution du blâme est engagé, les observateurs passent en revue les faits disponibles de manière biaisée en exagérant le contrôle volitif ou causal de l'agent et en minimisant leurs critères habituels d'attribution du blâme, ou encore en sélectionnant et recherchant les faits qui peuvent concourir à l'attribution de ce blâme. En sus de ces influences sur l'évaluation spontanée des faits, le processus d'attribution du blâme est également facilité par des facteurs tels que la tendance à la surestimation du contrôle qu'exercent les agents sur leurs actions et à confirmation des attentes négatives » (Alicke cité in T. NADELHOFFER, « Experimental Philosophy of Action », *New Waves in Philosophy of Action* (A. B. J. AGUILAR et K. FRANKISH dir.), New York, Palgrave MacMillan, 2011, pp. 50-78, p. 61.



« Parce que les participants pensent que le président de l'entreprise est un personnage digne de blâme qui a en toute connaissance de cause provoqué un effet secondaire nuisible, ils sont amenés à croire, de manière erronée, qu'il a intentionnellement nui à l'environnement. Dans le cas positif, au contraire, parce que les participants sont à nouveau devant un responsable d'entreprise qui ne manifeste aucun souci de l'environnement, ils répugnent à juger qu'il a pu contribuer à une amélioration écologique, car cela suggérerait par ailleurs qu'ils apprécient ce responsable. Dans les deux cas, les participants ne font pas tant un usage approprié du concept d'action intentionnelle qu'un usage détourné, dans leur désir de blâmer le président pour son attitude désinvolte à l'égard de l'environnement. » (380)

190. En résumé, Nadelhoffer suggère que les intuitions relatives à la culpabilité de l'agent sont déclenchées par des processus inconscients affectifs et qu'elles sous-tendent (et biaisent de manière inappropriée) les attributions asymétriques d'intentionnalité des sujets naïfs dans le cas des scénarios de Joshua Knobe. Le président est considéré comme coupable dans les *deux* scénarios parce qu'il ne se préoccupe pas des conséquences environnementales de son action. L'asymétrie n'est donc pas liée à la valence morale de l'effet secondaire lui-même, mais à notre façon habituelle d'appréhender les individus que nous estimons coupables ou blâmables. Dans la version négative du scénario de Joshua Knobe, le fait que le président est jugé digne de blâme donne libre cours à l'attribution d'une intentionnalité à son action. Dans la version positive, c'est l'inverse, l'attribution d'intentionnalité n'est évidemment pas induite par l'attitude de blâme sous-jacente.

SECTION 2 – EFFET KNOBE ET JURISPRUDENCE

191. Examinons à présent l'effet Knobe, selon les diverses explications qu'il a reçues, dans le contexte de la jurisprudence. Etant donné l'accumulation de données sur ce phénomène issues d'intuitions naïves de sujets profanes, il nous semble que, du point de vue de la pratique judiciaire, le problème pourra intéresser plus spécialement les systèmes pénaux impliquant des jurys populaires, comme c'est notamment le cas aux États-Unis, en Grande-Bretagne, ou en France dans les jugements

(380) T. NADELHOFFER, « Experimental Philosophy of Action », *op. cit.*, p. 62.



d'assises. Comme le souligne Nadelhoffer, l'un des rares philosophes à avoir souligné la pertinence de l'effet Knobe pour le droit :

« Si l'immoralité d'une action ou d'un effet secondaire tend à biaiser les attributions ordinaires d'intentionnalité, tous les crimes majeurs, tels que les meurtres ou les viols sont immoraux en plus d'être illégaux, et par conséquent la capacité d'un juré à déterminer la *mens rea* d'un prévenu de manière non biaisée est sérieusement entachée. » (381)

192. Les jurés de procès de meurtre au premier degré aux États-Unis sont habituellement informés que l'accusé peut être déclaré coupable s'il a causé la mort de la victime (c'est-à-dire : s'il est l'auteur d'un *actus reus* ou « acte coupable ») et si le meurtre était illégal, volontaire, délibéré et prémédité (c'est-à-dire : s'il a commis son acte avec la *mens rea* appropriée ou un état d'esprit coupable). Les instructions type données aux jurys criminels dans le Vermont (*Vermont Model Criminal Jury Instructions*), pour ne prendre que cet exemple, glosent « volontaire » (*willful*) de la manière suivante :

« Un acte volontaire est un acte qui est réalisé en connaissance de cause et de manière intentionnelle. Vous estimerez que le meurtre est volontaire si [le prévenu] a tué [la victime] exprès, et non pas accidentellement, par inadvertance ou non intentionnellement » (382).

193. De manière cruciale, l'*actus reus* et la *mens rea* (383) doivent être estimés indépendamment l'un de l'autre. À défaut de quoi la distinction entre un accident et un crime entraînant le même dommage ne saurait être maintenue. L'effet Knobe met précisément l'accent sur la difficulté à séparer mentalement ces deux facteurs au cœur des jugements de culpabilité. Il le fait de la manière suivante : la question de la culpabilité et du degré de culpabilité d'un accusé suspecté d'un crime est réglée lorsque l'on peut déterminer si ce dernier a agi avec la *mens rea* appropriée, laquelle doit être évaluée indépendamment de l'*actus reus*. Mais, du fait de l'effet Knobe, certains aspects de l'*actus reus*, en particulier ses propriétés morales et la sévérité des conséquences qu'il entraîne,

(381) T. NADELHOFFER, « Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality », *op. cit.*, p. 204.

(382) *Vermont Model Jury Instructions on first degree murder* (2006), 13 V.S.A. § 2301, CR24-031.

(383) La *mens reas*, du moins dans les systèmes de droit anglo-saxons, tend elle-même à être répartie en différentes sous-catégories (intentionnalité au sens standard, connaissance des conséquences de l'action, insouciance, négligence).

semblent bien influencer l'attribution d'une *mens rea*. La division du travail, censément claire, entre l'*actus reus* et la *mens rea*, perd alors toute sa consistance et avec elle la distinction cruciale entre des accidents dommageables et des crimes intentionnels. Du point de vue de l'accusé cela signifie que la gravité des conséquences de son action accroît la probabilité qu'on lui attribue une intention coupable.

194. Ce risque est d'autant plus grand que l'on tend à considérer l'effet Knobe comme un biais psychologique, comme le fait, par exemple, Nadelhoffer. Dans son interprétation de l'effet Knobe, le processus d'attribution de la *mens rea*, de l'intention coupable, subit clairement une distorsion : des processus subconscients et émotionnels d'un sentiment de blâme ou d'accusation sous-tendent à leur tour les attributions asymétriques d'intentionnalité. Dans une approche comme celle de Joshua Knobe, selon laquelle la valence morale, plutôt que le sentiment de blâme ou de mise en accusation, expliquerait l'effet qu'il a mis en évidence, le problème de la partialité du jury, tel que Nadelhoffer le soulève, ne serait toutefois qu'à moitié résolu :

« Après tout, si les jurés d'un procès concernant un meurtre sordide sont davantage susceptibles de juger que le prévenu a causé la mort de la victime de manière intentionnelle parce que cette mort est perçue comme étant intrinsèquement un mal, alors, dans la plupart des cas impliquant des crimes sérieux, les dés sont clairement pipés en défaveur des prévenus dès le début de la partie. Ce problème persiste même si la culpabilité du prévenu n'affecte pas de la sorte les jugements d'intentionnalité émis par les jurés. » (384)

195. Ces implications pourraient être en principe atténuées dans des systèmes légaux dans lesquels l'appel à des jurys non professionnels n'a pas cours. Les experts légaux – avocats, spécialistes universitaires du droit, juges – peut-on supposer – doivent posséder et manier un concept d'intentionnalité plus sophistiqué et plus adéquat, plus en accord avec les faits qu'on leur présente, étant donné que ce concept siège au plus profond de leur discipline. Autrement dit, ces professionnels du droit, même si leurs métiers présentent une certaine diversité, doivent en principe user d'un concept d'intentionnalité cohérent et qui s'accorde avec sa définition par la théorie de l'action et la philosophie légale. Si cette hypothèse de cohérence entre la théorie et la pratique

(384) T. NADELHOFFER, « Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality », *op. cit.*, p. 215.

professionnelle a du sens, et si, plus précisément, la prescription de l'usage d'un concept d'intentionnalité insensible à des considérations morales adjacentes est jugée recevable, nous disposons alors d'un argument assez fort en faveur de jurys uniquement composé d'experts et de professionnels plutôt que de profanes issus du public. Dans la suite de cet article, nous allons présenter des résultats expérimentaux qui offrent un éclairage sur cette hypothèse et ses conclusions possibles.

SECTION 3 – ATTRIBUTIONS D'INTENTIONNALITÉ CHEZ DES PROFESSIONNELS FRANÇAIS DU DROIT

1. PARTICIPANTS

196. 71 juristes professionnels français, des professeurs de droit et des doctorants, ont été invités à remplir un questionnaire en ligne. 35 participants étaient de sexe féminin. Les non-locuteurs maternels du français et les participants qui répondaient au premier scénario en moins de vingt secondes ont été éliminés. Il restait 59 fichiers de données valides, parmi lesquels 29 remplis par des participants de sexe féminin.

2. MÉTHODE

197. Notre questionnaire a été conçu et administré à l'aide du logiciel en ligne Qualtrics. On présentait aux participants de manière aléatoire la version négative ou bien la version positive du scénario environnemental de Joshua Knobe, ici de nouveau reproduit (la version positive est entre crochets) :

Le vice-président d'une entreprise va voir le président et dit : « Nous avons mis au point un nouveau programme. Celui-ci nous permettra d'augmenter considérablement nos profits mais [et] il aura aussi pour effet de nuire à [d'aider] l'environnement ». Le président répond : « Je me fiche complètement de nuire à [d'aider] l'environnement. Tout ce qui m'intéresse, c'est de faire le plus de profits possibles. Démarrons ce nouveau programme ». Le nouveau programme est démarré. Comme prévu, l'environnement en pâtit [bénéficie]. »

198. Les participants sont invités à répondre par oui ou par non (choix forcé) à la question portant sur le fait de savoir si le président a nui ou aidé l'environnement de manière intentionnelle. Ensuite, ils devaient déclarer dans quelle mesure ils étaient d'accord avec l'énoncé « le président a nui/aidé l'environnement de manière intentionnel » sur une échelle de Likert à 7 points allant de (1) 'fortement en désaccord' à (7) "tout à fait d'accord", avec au point 4 de l'échelle la mention 'ni d'accord, ni en désaccord'. Enfin, on demandait aux participants s'ils pensaient que le président était digne de louange ou de blâme, ou de ni l'un ni l'autre, pour son action.

199. À la suite de ces questions relatives à la première version du scénario, on présentait aux participants l'autre version du scénario. S'ils avaient reçu la vignette positive d'abord, ils devaient à présent considérer la vignette négative, et inversement. Le choix de ce design expérimental intra-sujets était motivé par les résultats de Shaun Nichols et Joseph Ulatowski discutés plus haut (385), selon lesquels on doit prendre en compte plusieurs interprétations possibles de l'expression « intentionnellement ». L'idée de notre design était donc de vérifier si l'application du concept d'intentionnalité par les juristes professionnels demeurerait cohérente entre les deux versions du scénario.

3. RÉSULTATS ET DISCUSSION

200. Les réponses à choix forcé différaient significativement entre les deux groupes (nuisance vs. aide). Dans l'ensemble, 59 % des participants ont procédé à une attribution d'intentionnalité dans le cas négatif (nuisance) et seulement 21 % dans le cas positif (aide à l'environnement) (386). Les résultats sont présentés dans la Figure 1.

(385) S. NICHOLS et J. ULATOWSKI, « Intuitions and Individual Differences: The Knobe Effect Revisited », *op. cit.*

(386) Test McNemar pour N=59, p<.001.

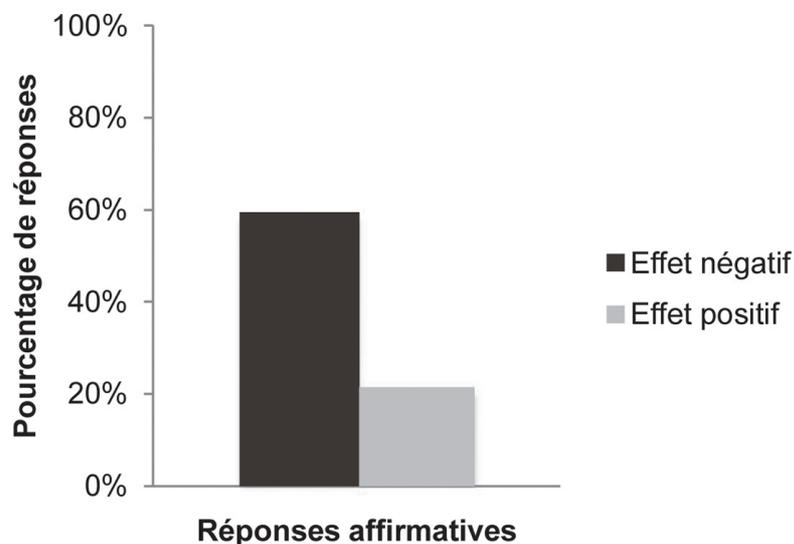


Figure 1. Pourcentage de réponses affirmatives, version négative *vs.* version positive.

201. Sur une échelle de Likert à 7 points, l'accord en moyenne avec l'affirmation que le président de l'entreprise a nui intentionnellement à l'environnement était de 4.64 (SD=2.11) ; et l'accord en moyenne avec l'affirmation que le président a aidé l'environnement était de 3.17 (SD=1.86), ce qui présente une différence significative (387), cf. figure . De manière encore plus significative, 59 % des participants ne donnaient pas la même réponse dans les deux cas. Si l'on reproduit le *design* inter-sujets habituellement utilisé par les études de l'effet Knobe, on accroît naturellement l'asymétrie des réponses. Dans ce cas, l'accord pour la version négative est en moyenne de 4.78 (SD=1.98), et pour la version positive de 2.72 (SD=1.67), une différence à nouveau significative (388).

(387) Échantillons couplés, $t(58)=5.768$, $p<.001$.

(388) Échantillons indépendants, $t(57)=-2.532$, $p=.014$.

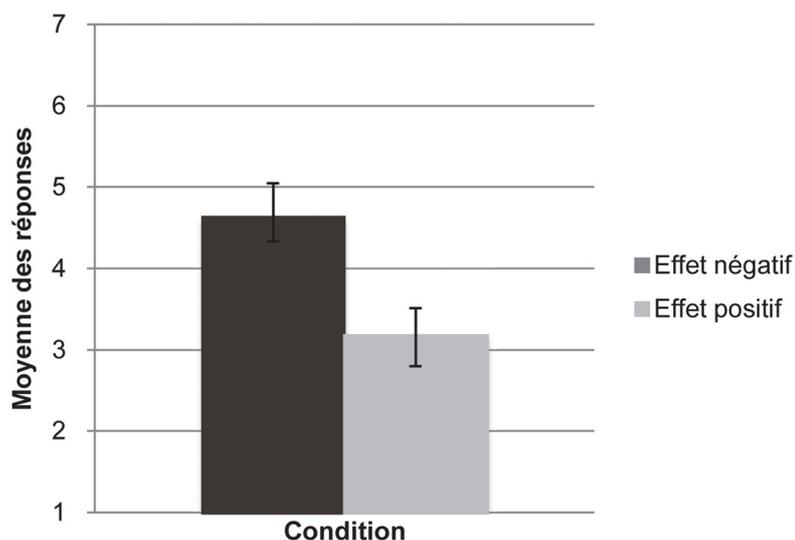


Figure 2. Assentiment moyen (les barres d'erreur indiquent l'erreur type à la moyenne).

202. Ces résultats sont plutôt clairs, qu'ils proviennent d'une élicitation des réponses par choix forcés ou par le moyen d'une échelle de Likert 1-7. Les intuitions des professionnels du droit que nous avons testés révèlent chez eux la présence d'un effet Knobe significatif. Le plus notable est que, selon notre design expérimental particulier, près de deux tiers des participants notaient différemment sur l'échelle de Likert les deux scénarios. Il semble donc que la conception de l'intentionnalité qui domine chez ces professionnels du droit est sensible à la valence morale des conséquences de l'action.

SECTION 4 – ATTRIBUTIONS D'INTENTIONNALITÉ CHEZ DES JUGES FRANÇAIS PROFESSIONNELS

203. Dans une autre étude, les auteurs de cet article, ont sondé les intuitions de juges français professionnels. Dans une première expérience inter-sujets qui soumettait aux participants les deux conditions du scénario de Joshua Knobe dans un ordre randomisé, l'effet habituel était reproduit. Pour la réponse à choix forcé [oui/non], 86 % des sujets faisaient une attribution d'intentionnalité dans le cas négatif, et

seulement 11 % dans le cas positif. (389) Il se trouve que ces résultats sont quasiment identiques avec les résultats originaux de Joshua Knobe (2003) obtenus sur des sujets profanes. 78 % fournissaient des données différentes, et peut-on dire contradictoires, pour les deux cas. Les réponses graduées sur l'échelle de Likert 1-7 démontraient également un effet Knobe important. L'accord avec l'affirmation que l'action du président était intentionnelle, dans le cas de sa conséquence négative, était de 5.67 sur 7 (SD=1.45), et de 2.56 sur 7 (SD=1.48) dans le cas positif. (390)

204. Un contre-argument possible quant à la portée de tels résultats est le suivant. Quand bien même des juges professionnels succomberaient à l'effet Knobe, cela n'aurait guère de conséquence quant à la justesse de leur jugement, dans la mesure où seuls les cas dont les conséquences sont négativement connotées ont une importance – on va rarement au tribunal pour avoir commis une bonne action, involontairement ou non. Bien qu'il puisse en effet y avoir une asymétrie dans l'attribution d'une intentionnalité à des actions dont les conséquences présentent des valences morales opposées, cette asymétrie n'est nullement incompatible avec une évaluation homogène des cas négatifs eux-mêmes, les seuls dont l'application de la loi a à se soucier. Mais nous rejetons cet argument. D'abord parce que les résultats négatifs et positifs d'une action ont tous deux autant d'importance dans les contextes légaux, lorsque, par exemple, des circonstances atténuantes sont en jeu. Ensuite, et à l'inverse de beaucoup de commentateurs, nous cherchons à saisir l'effet Knobe à sa racine, et non comme une simple asymétrie dans l'attribution d'intentionnalité à des *types* d'actions (positives ou négatives). Nous voudrions en particulier faire ressortir le caractère *graduel* de cet effet, selon une corrélation positive peut s'établir entre le degré de dommage envisagé comme conséquence de l'action incriminée et la propension avec laquelle les individus considèrent cette action comme forcément intentionnelle. Pour résumer cette approche en une simple formule, nous suspectons que plus la conséquence est dommageable, plus probable est l'attribution d'une *mens rea*.

205. Nous avons ainsi testé dans quelle mesure des juges français professionnels tendent à attribuer le caractère d'intentionnalité en présence de deux cas négatifs, l'un présentant une conséquence assez mauvaise, mais réparable, et l'autre une conséquence très mauvaise et

(389) Test McNemar exact test pour N=36, p<.001.

(390) Échantillons couplés, t(35)= 8.383, p<.001, 95 % CI [2.358;3.865].



irréversible (391). Dans le cas très négatif, le niveau d'attribution de l'intentionnalité était de 5.18 (SD=1.74) sur une échelle de Likert à 7 points, ce qui différait significativement du cas moins négatif (M=3.33 ; SD=1.76) (392). Ceci tend à montrer que l'effet Knobe est mieux saisi si on en adopte une lecture graduelle, plutôt qu'une interprétation plus binaire prédisant des attributions d'intentionnalité opposées selon la valence morale des conséquences envisagées (393). La conséquence, pour le droit, semble être que plus le résultat de l'*actus reus* sera négatif, plus grande sera la probabilité que l'accusé se voie attribuer la *mens rea* correspondante.

206. Ces nouveaux résultats montrent que les professionnels du droit, du moins ceux que nous avons eu l'opportunité de tester en France, sont fortement sensibles à l'effet Knobe. Le concept de *mens rea* mis en œuvre par ces professionnels est sensible, pourrait-on dire, à la désirabilité des conséquences de l'acte. Comme nous venons de le voir, cette sensibilité ne revêt pas une nature bivalente simple (attribution d'intentionnalité pour les actes à conséquences négatives, mais pas pour les actes à conséquences positives), comme on le suppose généralement dans la littérature secondaire. Différemment, ces juges, avocats ou professeurs de droit, comme le tout-venant, opèrent avec un concept d'intentionnalité qui n'est pas simplement variable selon que l'action engendre un effet moralement valorisant ou moralement néfaste, mais selon que cet effet est plus ou moins indésirable.

207. Que pouvons-nous en conclure ? Cela dépend de l'interprétation de l'effet Knobe que l'on tend à considérer comme correcte. Si, suivant Joshua Knobe lui-même, on estime que le concept d'intentionnalité, comme la plupart des concepts psychologiques ordinaires, est sensible aux considérations morales, et que les humains sont des « moralisateurs impénitents », une attribution asymétrique du concept d'intentionnalité dans le contexte des scénarios de Joshua Knobe n'est nullement incompatible avec l'idée qu'une telle attribution est parfaitement compétente. Ou, autrement dit, si l'on adopte l'explication du *side-effect effect* en termes de compétence, nos résultats, du moins de prime abord, ne sont ni surprenants ni inquiétants.

(391) M. KNEER et S. BOURGEOIS-GIRONDE, « Guilty Minds and Biased Minds », manuscrit.

(392) Échantillons indépendants, $t(30)=-2.974$, $p=.006$, 95 % CI [-3.109 ; -0.578], test de Cohen $d=1.06$.

(393) Kneer rapporte des données selon lesquelles l'interprétation graduelle de l'effet Knobe est celle qui correspond le mieux aux attributions d'intentionnalité opérées par les profanes (M. KNEER, « Beyond Bivalence: A Graded Model of the Knobe Effect? », *op. cit.*).



208. Si, en revanche, on pense que la sensibilité morale du *side-effect effect* est un biais qui provoque une distorsion des jugements et les rend inadéquats, nos résultats peuvent être source de préoccupation. Dans les systèmes judiciaires où les jurys populaires n'ont pas cours, le fait que le profane ait une conception biaisée de l'intentionnalité ne prête pas à conséquence. Mais si les juristes professionnels, et en particulier les juges, témoignent de la même conception, les conséquences ne sont plus vraiment négligeables. Dans la mesure où les crimes, en sus d'être illégaux, sont des atteintes immorales, et dans la mesure où les considérations morales distordent les attributions d'intentionnalité, il semble impossible que les juges professionnels évaluent la *mens rea* d'un prévenu de manière indemne aux biais. En d'autres termes, du fait de la prégnance de ce biais psychologique, la jurisprudence nous semble être confrontée à un sérieux problème. L'évaluation de la culpabilité juridique dépend, en partie, de l'évaluation indépendante d'une *mens rea* appropriée, indépendante, plus précisément, d'aspects d'un *actus reus* tels que la gravité de ses conséquences ou la dimension morale de ces dernières. Or, il s'avère que ces attributions de la *mens rea* sont nettement sensibles à certaines caractéristiques de l'*actus reus*, et, significativement, du degré de désirabilité des conséquences.

209. Nous avons indiqué plus haut que nos résultats, en première approche, n'affectaient pas les interprètes de l'effet Knobe en termes de compétence. L'application du concept d'intentionnalité par les professionnels du droit n'est pas mise en cause si l'on pense qu'ils appliquent en fait et doivent appliquer, selon cette approche en termes de compétences, un concept moralement connoté. Cependant, le concept théorique d'intentionnalité, tel qu'il est censé prévaloir dans la théorie du droit et dans l'exercice du droit lui-même, n'autorise en principe pas une telle sensibilité aux caractéristiques morales d'une action. À l'instar du concept dominant en philosophie du droit et en théorie de l'action, le concept d'intentionnalité est, dans ce contexte théorique, envisagé indépendamment de préoccupations morales et n'autorise donc pas une attribution d'intentionnalité dépendante de caractéristiques morales. Cela signifie que, bien que d'un côté le concept d'intentionnalité appliqué par les praticiens du droit puisse être compatible avec certaines théories psychologiques récentes, il n'en demeure pas moins, d'un autre côté, en porte à faux avec la théorie et les principes de la pratique du droit lui-même. Si ce que requièrent ces principes ne se retrouve pas en pratique, il y a peut-être un problème fondamental, sur les relations psychologiques implicites, et non explicites, car celles-ci sont claires,

à analyser sur la relation entre droit et intuitions morales. C'est ce que nous avons cherché à faire dans notre étude.

210. Il nous faut préciser un dernier point eu égard aux “interprétations multiples” qui ont été données de l'effet Knobe et du problème de l'attribution ordinaire du concept d'intentionnalité. Comme nous l'avons indiqué plus haut, des auteurs tels que Shaun Nichols et Joseph Ulatowski ou Alessandro Lanteri ont mis en évidence la possibilité qu'il n'y ait pas un concept simple et unifié d'intentionnalité, et par conséquent pas de critères observables uniques correspondant à son application. Les individus qui donnent des réponses similaires aux deux versions du scénario (affirmatives ou négatives) et ceux qui modifient leurs réponses selon les deux versions, peuvent très bien fonctionner avec trois concepts d'intentionnalité distincts, tous trois légitimes. Nous aurions peut-être ici un motif pour nous départir du présupposé théorique, possiblement empiriquement mal fondé, selon lequel le concept d'intentionnalité est irrécusablement uniforme. Toutefois, comme nous le montrent les résultats, il y a bien une interprétation et une application dominantes prévalant chez les juristes du concept d'intentionnalité. Un concept d'intentionnalité, de fait contradictoire selon le contexte auquel il s'applique, et sensible aux dimensions morales des actes envisagés, semble prévaloir auprès d'environ deux tiers des individus interrogés ; alors que la désambiguïsation de concept en deux concepts cohérents ne peut rendre compte que d'un tiers de l'ensemble des réponses que nous avons obtenues.

211. En réalité, peu importe que cette conception soit uniforme ou simplement fortement dominante, il ne fait pas de doute que les attributions d'intentionnalité qu'elle sous-tend ne sont a priori pas conciliables avec le concept théorique qui prévaut en droit. De plus, bien que la pluralité des interprétations ou la structure ouverte d'un concept fasse clairement sens quand on étudie la psychologie ordinaire, il est moins évident qu'elles soient si facilement acceptables en théorie du droit. Il nous semble difficile d'imaginer une pratique juridique cohérente qui s'appuierait sur trois (ou plus) conceptions différentes de l'action intentionnelle. Aussi, les arguments qui peuvent être convoqués pour rendre compte de la mise en œuvre ordinaire d'intuitions morales perdent assez vite de leur pouvoir justificatif quand on vient à s'intéresser au rôle de ces mêmes intuitions dans un contexte juridique.