

UNE REVUE MULTIDISCIPLINAIRE SUR LES ENJEUX NORMATIFS
DES POLITIQUES PUBLIQUES ET DES PRATIQUES SOCIALES.

Les ateliers de l'éthique The Ethics Forum

A MULTIDISCIPLINARY JOURNAL ON THE NORMATIVE
CHALLENGES OF PUBLIC POLICIES AND SOCIAL PRACTICES.

- 4-30 Volitions, préoccupations prudentielles, et réflexivité dans la psychopathie **Mathieu Garcia**
- 31-45 Is Daniel A Monster? Reflections on Daniel A. Bell and Wang Pei's "Subordination Without Cruelty" Thesis
Rainer Ebert, Valéry Giroux, Angie Pepper, Kristin Voigt
- 46-71 Justice, Impartiality, and Equality in the Allocation of Scarce Vaccines: A Reply to Saunders **Hugh V. McLachlan**

DOSSIER :

Affective States, Happiness, and Well-Being

Sous la direction de Samuel Lepine

- 72-80 Introduction **Samuel Lepine**
- 81-111 Happiness and the Metaphysics of Affect **Daniel M. Haybron**
- 112-125 Une théorie du bien-être comme bonheur approprié **Mauro Rossi, Christine Tappolet**
- 126-150 Virtue, Happiness, and Emotion **Antti Kauppinen**
- 151-175 Les bonnes humeurs, le bonheur et le bien-être **Samuel Lepine**
- 176-191 The Hedonist's Emotions **Julien Deonna, Fabrice Teroni**
- 192-213 Peut-on encore être hédoniste éthique ? **Antonin Broi**
- 214-241 Quelle pro-attitude pour le subjectivisme à propos du bien-être ? **Stéphane Lemaire**
- 242-265 bien-être, affectivité et société : enjeux moraux et enjeux structureaux **Julien Claparède-Petitpierre**

UNE REVUE MULTIDISCIPLINAIRE SUR LES ENJEUX NORMATIFS
DES POLITIQUES PUBLIQUES ET DES PRATIQUES SOCIALES.

Les ateliers de l'éthique The Ethics Forum

A MULTIDISCIPLINARY JOURNAL ON THE NORMATIVE CHALLENGES
OF PUBLIC POLICIES AND SOCIAL PRACTICES.

La revue est financée par le Conseil de recherches en sciences humaines (CRSH)
et administrée par le Centre de recherche en éthique (CRE)

The journal is funded by the Social Sciences and Humanities Research Council (SSHRC)
and administered by the Center for Research on Ethics (CRE)



Attribution 4.0 International

Vous êtes libres de reproduire, distribuer et communiquer les textes de cette revue au public, transformer et créer à partir du matériel pour toute utilisation y compris commerciale, selon les conditions suivantes :

- Vous devez créditer l'Œuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l'Œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son Œuvre.

Pour tous les détails, veuillez vous référer à l'adresse suivante :
<https://creativecommons.org/licenses/by/4.0/legalcode>

You are free to copy and distribute all texts of this journal, transform, and build upon the material for any purpose, even commercially, under the following conditions:

- You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

For all details please refer to the following address:
<https://creativecommons.org/licenses/by/4.0/legalcode>

RÉDACTRICES EN CHEF/EDITOR

Ryoa Chung, Université de Montréal

Kristin Voigt, Université McGill

COORDONNATEUR DE RÉDACTION/ADMINISTRATIVE EDITOR

Éliot Litalien, Université de Montréal

COMITÉ EXÉCUTIF DE RÉDACTEURS/EXECUTIVE EDITORS

Éthique fondamentale : Natalie Stoljar, Université McGill

Éthique et politique : Vacant

Éthique et santé : Vacant

Éthique et économie : Vacant

Éthique et environnement : Vacant

COMITÉ D'EXPERTS/BOARD OF REFEREES:

Arash Abizadeh, Université McGill

Charles Blattberg, Université de Montréal

Marc-Antoine Dilhac, Université de Montréal

Francis Dupuis-Déri, Université du Québec à Montréal

Geneviève Fuji Johnson, Université Simon Fraser

Axel Gosseries, Université catholique de Louvain

Joseph Heath, Université de Toronto

Samia A. Hurst, Université de Genève

Julie Lavigne, Université du Québec à Montréal

Robert Leckey, Université McGill

Catriona Mackenzie, Macquarie University

Katie McShane, Colorado State University

Bruce Maxwell, UQTR

Pierre-Yves Néron, Université catholique de Lille

Wayne Norman, Duke University

Hervé Pourtois, Université catholique de Louvain

Vardit Ravitsky, Université de Montréal

Mauro Rossi, Université du Québec à Montréal

Christine Straehle, Université d'Ottawa

Daniel M. Weinstock, Université McGill

Jennifer Welchman, University of Alberta

NOTE AUX AUTEURS

Merci de soumettre votre article à l'adresse :
eliot.litalien@umontreal.ca.

Les articles seront évalués de manière anonyme par deux pairs. Les consignes aux auteurs se retrouvent sur le site de la revue (www.lecre.umontreal.ca/categorie/ateliers_ethique_ethics_forum/). Tout article ne s'y conformant pas sera automatiquement refusé.

GUIDELINES FOR AUTHORS

Please submit your paper to: eliot.litalien@umontreal.ca.

Articles are anonymously peer-reviewed. Instructions to authors are available on the journal website (www.lecre.umontreal.ca/categorie/ateliers_ethique_ethics_forum/). Papers not following these will be automatically rejected.

VOLITIONS, PRÉOCCUPATIONS PRUDENTIELLES, ET RÉFLEXIVITÉ DANS LA PSYCHOPATHIE

MATHIEU GARCIA
UNIVERSITÉ DE BORDEAUX

RÉSUMÉ :

Conformément à une tradition faisant des sujets manifestant un fonctionnement psychopathique des êtres *moralelement aveugles*, voire *moralelement morts*, les discussions ayant trait à leur agentivité morale ont pendant longtemps jugé subsidiaire l'analyse des mécanismes volitifs et conatifs, pourtant essentielle pour comprendre leurs agissements immoraux. Partant plus spécifiquement du décalage qui semble, au regard de certaines données récemment publiées, s'opérer chez ceux présentant ce type de personnalité entre jugements moraux et choix moraux, nous proposons dans cet article de discuter tour à tour les théories susceptibles de venir expliquer ce fossé qui sépare ce que le psychopathe reconnaît (ou peut, d'une manière générale, reconnaître) et ce qu'il fait. Après avoir présenté l'hypothèse – problématique en plus d'un sens – en faveur d'une faiblesse de la volonté associant la psychopathie à l'acrasie, nous exposons les apports respectifs des conceptions rationalistes et néosentimentalistes, avant d'interroger le présupposé qu'elles partagent avec les théories de l'incontinence psychopathique. Nous argumentons en effet que chacune de ces trois positions tend implicitement à octroyer à ces individus dysso-ciaux une propension à recourir à une certaine forme d'activité délibérative peu étayée sur le plan empirique. Une discussion resserrée autour des caractéristiques prudentielles relevées dans la psychopathie nous conduira au contraire à défendre l'idée selon laquelle l'immoralité de ceux qui en sont atteints n'est, dans la plupart des cas, qu'un effet collatéral d'un manque d'appétence pour le processus d'autoévaluation réflexive. Et nous verrons qu'il y a de fortes raisons de croire que cette irréflexivité – qui rend compte d'un grand nombre de facettes du mode d'existence de la conscience psychopathique tout en restant compatible avec les principaux modèles explicatifs aujourd'hui privilégiés pour éclairer les décisions morales qui s'en dégagent – est directement liée à la vie affective (extrêmement pauvre) du psychopathe. Nous terminons en avançant que ce lien entre déficit émotionnel et carence réflexive fait apparaître de nouvelles perspectives pour penser les différents types de profils moraux retrouvés au sein des diverses formes de psychopathie que nous suggérons de répartir le long d'un continuum sur lequel s'étendent plusieurs niveaux (ou degrés) de déficience expérientielle.

ABSTRACT:

In accordance with a tradition depicting individuals with psychopathic functioning as *morally blind*, or even *morally dead*, discussions concerning their moral agency have for a long time neglected the analysis of volitional and conative aspects, whose contribution is nevertheless crucial in our understanding of their immoral actions. Based on the findings about the discrepancy that seems to occur, in this type of personality, between moral judgments and moral choices, we propose in this article to discuss the several theories that might explain this gap between what the psychopath recognizes (or can, in a general way, recognizes) and what he does. After having presented the hypothesis - problematic in more than one sense - of a weakness of will associating psychopathy with acrasia, we expose the respective contributions of rationalist and neo-sentimentalist conceptions, before questioning the presupposition they share with theories of psychopathic incontinence. We argue that each of these three positions implicitly tends, with little empirical support, to ascribe to these dyssocial individuals a propensity to resort to some form of deliberative activity. An in-depth discussion of the prudential characteristics of psychopathy will lead us, on the contrary, to defend the idea that the immorality of those who suffer from it is, in most cases, only a collateral effect of a lack of appetite for the process of reflective self-evaluation. And we will see that there are strong reasons to believe that this lack of reflection - which accounts for many facets of the psychopathy construct while remaining compatible with the main explanatory models now favored to understand the psychopath's moral decisions - is directly related to the (extremely poor) emotional life of psychopaths. We conclude by arguing that this way of linking emotional deficit and reflective impairment offers new perspectives for thinking about the different types of moral profiles found within the various forms of psychopathy that we suggest to distribute along a continuum on which levels of experiential deficiency extend.

INTRODUCTION

La psychopathie est couramment décrite comme un trouble de la personnalité comprenant une constellation de traits opérationnalisés en trois ou quatre dimensions (appelés aussi « facteurs »). On distingue ainsi généralement une dimension interpersonnelle (manipulation, soi grandiose, mensonge pathologique, charme superficiel, etc.), une dimension affective (manque de remords et d'empathie, détachement et insensibilité émotionnels, froideur affective, affects superficiels, etc.), une dimension comportementale ou liée au style de vie (impulsivité, recherche de stimulation, conduites à risque, objectifs de vie peu réalistes, présentisme, etc.) et une dimension antisociale (conduites délictueuses et criminelles, tendances agressives et antisociales, troubles du comportement précoces, etc.) (Andershed, 2010; Cooke et Michie, 2001; Hare, 2003; Hare et Neumann, 2006; 2008; Harris, Skilling et Rice, 2001; Patrick, 2018). Les individus présentant ces caractéristiques psychopathiques sont non seulement connus pour leur manque d'empathie (notamment affective), mais également, et corrélativement, pour leur manque de considération des autres êtres humains, lesquels se trouvent être régulièrement objectalisés et instrumentalisés (McIlwain, 2010; Baron-Cohen, 2012). Ils s'avèrent en outre le plus souvent peu touchés – d'un point de vue émotionnel – par la confrontation à des situations/stimuli aversifs et/ou désagréables (p. ex. images d'enfants affamés, de corps mutilés, etc.) (Anastasiou-Hadjicharalambous et Warden, 2008; Blair, Colledge, Murray et Mitchell, 2001; Marsh *et al.*, 2013; De Wied, van Boxtel, Matthys et Meeus, 2012). Cet émoussement affectif – corroboré par nombre de travaux en neurophysiologie faisant état d'une relation étroite entre psychopathie et sous-activation amygdalienne (Cheng, Hung et Decety, 2012; Syngelaki, Fairchild, Moore, Savage et van Goozen, 2013) – apparaît d'autre part renforcé, ou peut-être sous-tendu, par d'importantes difficultés démontrées à identifier/reconnaître, traiter et comprendre les stimuli de nature affective, comme la détresse, la peur ou encore la tristesse (Marsh et Blair, 2008; De Wied *et al.* 2012; Hodsoll, Lavie et Viding, 2014). Le décodage des expressions faciales d'autrui et des émotions qu'elles expriment constitue, en soi, une tâche ardue pour les personnes présentant des niveaux élevés de traits psychopathiques (voir Marsh et Blair, 2008). Notons enfin qu'à ces spécificités cognitives et émotionnelles se greffent des taux de criminalité et d'agressivité particulièrement élevés, ainsi qu'une propension moindre à s'engager dans des comportements prosociaux (Asscher *et al.*, 2011; Blair, 2010; Frick et Marsee, 2018; Patrick, 2018).

Qu'en est-il sur le plan strictement moral ? Pour répondre, on partira de l'ouvrage fondateur, *The Mask of Sanity*, publié pour la première fois en 1941, où l'auteur, Hervey Cleckley, postule la présence d'un *sens moral* significativement altéré comme composante centrale de la psychopathie. Bien avant lui, le médecin britannique James Prichard (1835) parlait de « folie morale » (*moral insanity*) pour qualifier ce qu'il considérait être une pathologie avant tout morale. De nombreux chercheurs ont pendant longtemps suivi cette voie et considéré les psychopathes comme des « monstres moraux » (*moral monsters*) (Ellis, 1890), ou des sujets « moralement morts » (Murphy, 1972), « moralement aveu-

gles » (*morally blind*) (Talbert, 2008), en raison de leur prédilection avérée pour des comportements contraires à l'éthique. Notons que ces descriptions classiques, insistant sur la prégnance d'une déficience morale (« *moral defectiveness* », « *moral imbecility* », « *moral insensibility* », Maudsley, 1873), s'accordent avec l'hypothèse – communément partagée (Furnham, Daoud et Swami, 2009) – selon laquelle les comportements antisociaux, délictuels et/ou criminels adoptés par ces individus résulteraient principalement d'une défaillance d'ordre moral. Et c'est précisément ce dysfonctionnement, touchant vraisemblablement à la fois les formes dites primaires et secondaires de la psychopathie, qui tendrait à expliquer les comportements immoraux pour lesquels optent ces individus (Duff, 1977; Leistico, Salekin, DeCoster et Rogers, 2008; Levy, 2014).

Néanmoins, il paraît difficile d'avancer que le psychopathe *opte* pour des comportements immoraux alors qu'il ne *sait* pas – étant donné son absence supposée (ou tout du moins son manque significatif) de sens moral – que ses comportements sont immoraux (de sorte qu'il vaudrait mieux considérer ces comportements comme étant *amoraux*). On retrouve ici le vieil adage socratique nous invitant à penser la personne au fonctionnement psychopathique comme n'étant pas « méchante volontairement », dans la mesure où le mal qu'elle commet découle tout simplement de son ignorance morale, d'une « incapacité morale » (Schramme, 2014) qui la conduit à ne pas *savoir* viser correctement le bien et à faire ainsi le mal tout en pensant faire le bien.

Est-il si clair, cependant, que la présence d'une personnalité psychopathique implique celle d'une incapacité morale ? C'est là la première question qui nous occupera. Et nous tâcherons de montrer, à l'appui des résultats empiriques les plus récents, qu'il n'est pas si évident de conclure à la présence d'un discernement moral aboli, ou même altéré, dans la psychopathie. Il s'agira en réalité de signaler l'existence d'un décalage quasi systématiquement observé dans ce type de fonctionnement entre jugements moraux et choix moraux. Ce premier résultat nous conduira à traiter une deuxième question, en l'occurrence celle de savoir ce qui peut expliquer ce décalage. Pour ce faire, nous présenterons et examinerons en détail les principales hypothèses concurrentes, défendues dans la littérature, destinées à en rendre compte. Nous procéderons en trois temps, qui traceront chacun une piste explicative. Nous nous attarderons d'abord sur les théories qui, en mettant l'accent sur la faiblesse de la volonté, font du psychopathe l'incarnation même du personnage acratique tel qu'a pu le dépeindre Aristote. Seront ensuite discutées les positions émotivistes (ou néosentimentalistes) attribuant un rôle prépondérant aux aspects émotionnels impliqués dans les comportements (im)moraux relevés dans cette pathologie de la personnalité. Nous poursuivrons notre exposé doxographique en évoquant les apports de l'approche cognitiviste (ou rationaliste) associant la psychopathie à une raison pratique défectueuse. Nous verrons cependant que chacune de ces trois voies d'explication (thèse de l'incontinence, de la cécité émotionnelle, et de l'irrationalité pratique), aussi intéressantes et justifiées soient-elles, repose sur un présupposé tenace consistant à doter le psychopathe d'un usage délibératif des connaissances morales qu'il semble posséder. Or, argumenterons-nous, cette

prédication d'une inclination à la délibération ne va pas de soi, et un certain nombre d'éléments cliniques convergent au contraire vers l'idée – que l'on trouve notamment chez Watson (2013) – selon laquelle les comportements immoraux du sujet aux traits psychopathiques résultent finalement en grande partie, et dans de nombreux cas, d'un déficit prudentiel que nous lierons, à rebours de certaines interprétations prioritairement (et plus ou moins explicitement) axées sur la faculté de voyage mental dans le temps, à une faible propension à l'autoévaluation réflexive. Nous soulignerons alors, par l'intermédiaire des travaux de Frankfurt (1971) sur la figure de l'irréflexif, en quoi cette thèse s'accorde non seulement avec l'ensemble des descriptions philosophiques et psychopathologiques de la psychopathie, et les différences facettes constitutives de ce trouble, mais également avec la quasi-totalité des théorisations disponibles pour éclaircir la ou les source(s) du divorce qui semble s'installer entre ce que le psychopathe reconnaît (ou sait) et ce qu'il fait (ou décide de faire). Nous terminerons néanmoins en soutenant, selon une logique renouant avec le paradigme néosentimentaliste, et dans un souci de concilier déficit émotionnel et déficit d'autoévaluation réflexive, que la seule façon d'expliquer ce manque, voire, parfois, cette absence totale, de réflexivité semble, en l'état actuel de nos connaissances, passer par une analyse resserrée sur les niveaux – dimensionnellement répartis le long du spectre des diverses expressions psychopathiques – d'altération de l'expérience émotionnelle.

QUELLES COMPÉTENCES MORALES POUR LE PSYCHOPATHE ?

Commençons par nous pencher sur la question du sens moral apparemment limité des sujets présentant une psychopathie. Plusieurs auteurs n'hésitent pas à souscrire à cette thèse qui suscitait déjà l'adhésion des aliénistes du XIX^e et des psychiatres français et anglo-saxons du XX^e. Rappelons à ce propos, pour bien cerner les enjeux, que pour un certain nombre de ces auteurs les bien nommés psychopathes mériteraient d'être pénalement irresponsabilisés, étant donné les difficultés qu'ils rencontrent à penser rationnellement les situations morales auxquelles ils peuvent se trouver confrontés (Morse, 2008). Levy (2008) avance par exemple que « les psychopathes ne possèdent pas les connaissances morales pertinentes » (p. 129) pour prétendre à une quelconque responsabilité morale (et donc, nous explique-t-on, pénale). Car sans ces connaissances, ils seraient « incapables de contrôler leurs actions à la lumière de raisons morales » (p. 129) éclairantes et dissuasives.

Or, cette conception est loin de faire l'unanimité, et devient même de plus en plus minoritaire. Les discussions font rage sur ce point, en particulier au sein de la communauté philosophique, où le concept de psychopathie nourrit plusieurs débats éthico-moraux qui ne sont, encore aujourd'hui, pas près d'être résolus. Cet intérêt semble en partie découler du fait que les individus présentant une personnalité psychopathique possèdent cette spécificité de manifester des déficits dont les répercussions stimulent et relancent certaines controverses situées à l'intersection de la philosophie et de la psychologie morale. Songeons notam-

ment à la place – de choix – que la notion de psychopathie occupe au sein des discussions métaéthiques opposant les rationalistes aux sentimentalistes moraux, les externalistes aux internalistes, ou encore les tenants de la compétence rationnelle à ceux de la compétence morale comme critères d’agentivité morale – sans oublier, bien entendu, les importantes questions, effleurées ci-dessus, qu’une telle configuration psycho(patho)logique soulève en termes de responsabilité morale *et* juridique.

En quel sens, donc, et sur quelles bases empiriques, peut-on être amenés à relativiser la position en faveur d’une psychopathie conçue sous l’angle de l’« aliénation morale » ? Eh bien, tout d’abord parce que les résultats jusqu’ici obtenus demeurent soit contradictoires, soit inconsistants (Borg et Sinnott-Armstrong, 2013). En témoigne la méta-analyse récemment publiée par Marshall et ses collaborateurs (2018), qui concluent à l’existence d’une relation extrêmement faible (bien que significative, certes) entre psychopathie et altération du jugement moral (quelles que soient les mesures et/ou les épreuves expérimentales utilisées), mais également entre psychopathie et pauvreté des raisonnements moraux déployés. Ces résultats viennent par conséquent nuancer la conception voulant que les psychopathes n’aient pas accès à certains principes moraux. D’autant que la même étude révèle par ailleurs que l’écrasante majorité des travaux disponibles à ce jour (aussi bien en psychologie qu’en neurosciences) indique que ces individus, loin d’être moralement aveugles, possèdent et investissent un certain nombre d’intuitions morales organisées autour des principes de loyauté et de respect.

En 2012, déjà, Aharoni et son équipe venaient récuser l’idée selon laquelle les personnes exhibant un fonctionnement psychopathique seraient dans l’incapacité de distinguer les transgressions (et les normes) morales des transgressions (et des normes) conventionnelles. Le score total de psychopathie ne prédit, en effet, aucunement la performance aux tâches de jugement proposées, ce qui vient contester le point de vue d’après lequel les psychopathes ne comprennent pas ce qui est moralement blâmable, et pour quelle(s) raison(s) ça l’est. Des résultats similaires sont rapportés à la suite d’études expérimentales examinant les jugements et les prises de décision face à des dilemmes moraux (Blair, 1995; Glenn *et al.*, 2009; Tassy *et al.*, 2009). La présence d’un fonctionnement psychopathique ne prédit, là encore, aucunement la teneur des jugements moraux émis. Et lorsque certaines de ces études mettent en évidence que la psychopathie est reliée à une probabilité accrue de réponses inconditionnellement utilitaristes (Glenn *et al.*, 2010; Bartels et Pizarro, 2011; Koenigs *et al.*, 2012), il ressort que cela n’est pas dû à une lacune inhérente aux concepts moraux manipulés ni même à un manque de compréhension des enjeux moraux sous-jacents aux scénarios et situations proposés, ou encore à l’insuffisante discrimination de ce qui est bien ou de ce qui est mal, mais bien plutôt au fait que ces jugements, aussi adéquats et justes qu’ils puissent être, ne les affectent guère et n’engendrent pas la réponse verbale ou comportementale correspondante (Cima *et al.*, 2010). En d’autres termes, et d’après ce qu’ont pu en dire les auteurs de ces études, les individus aux caractéristiques psychopathiques sauraient pertinemment ce qui est

moralement interdit ou approprié, mais ne se soucient pas de ce savoir moral. Ils n'en tiennent pas compte. Ce qui apparaît plutôt cohérent lorsque l'on sait que ces derniers présentent des niveaux standards – sinon élevés – d'intelligence, couplés à une certaine habileté quant aux raisonnements extra-moraux qu'ils peuvent mobiliser (Cleckley, 1941; Hare, 2003; Glenn *et al.*, 2010).

Partant de cet écart qui sépare l'intégrité des capacités de jugement moral des comportements moraux effectivement adoptés, plusieurs auteurs font l'hypothèse de l'existence de processus indépendants à l'origine du jugement *et* du comportement concrètement privilégié. C'est par exemple ce que révèle le cas des patients atteints de lésions cérébrales ventromédiales préfrontales, qui présentent des comportements sociaux inappropriés en dépit d'une capacité de jugement préservée (Eslinger et Damasio, 1985). Il paraît donc concevable, dans cette optique, de soutenir que le choix moral de l'action tel qu'il est reflété ou exprimé dans le comportement réel puisse être indépendant du jugement moral. De récentes recherches vont d'ailleurs dans le sens de cette apparente dissociation entre jugement moral et action morale, avec des preuves expérimentales en faveur d'une divergence – parfois massive – entre le jugement et le choix de l'action lors d'une évaluation morale (Kurzban *et al.*, 2012; Tassy *et al.*, 2013b). Le jugement ne présage pas du choix, et, réciproquement, le choix ne dit rien du jugement. Tassy et son équipe (2013a) ont pu à cet égard montrer qu'un niveau élevé de traits psychopathiques prédisait une plus grande proportion de réponses utilitaires pour le choix, mais pas pour le jugement moral. Ce qui pourrait expliquer les résultats contradictoires des études susmentionnées, où ces deux processus n'étaient pas différenciés et analysés de manière isolée.

On entrevoit donc, avec ces quelques travaux, que la présence, si fréquemment invoquée, d'un déficit moral, est loin d'aller de soi. Il se pourrait même que l'activité morale judicative du sujet psychopathe soit tout à fait opérante, sans pour autant que cela se traduise par des décisions morales pertinentes ou ajustées. Mais comment, dès lors, interpréter ce décalage entre jugement moral et choix moral ?

LE PSYCHOPATHE EST-IL ACRATIQUE ?

Un premier type de réponse pourrait consister à dire qu'il est la marque d'un conflit entre ce que l'appétit recherche et ce que la raison reconnaît comme le meilleur. Or, on le sait, cela ferait du psychopathe un acratique par excellence, puisque chez l'acratique, en effet, l'appétit lutte et s'oppose à la raison (Aristote, *Éthique à Nicomaque* [EN], I, 13, 1102b 17-18), en tant que leurs impulsions respectives vont en sens contraire (1102b 21). On devient acratique, écrit Aristote dans l'*Éthique à Eudème*, lorsque l'on passe à l'acte en fonction de l'appétit (II, 7, 1223a 38-39) contre ce que dicte la raison. L'appétit prend donc le dessus sur cette raison et sort vainqueur du combat. Il n'est donc pas question, ici, *a priori*, d'un déficit de rationalité, qui s'exprimerait à travers des jugements moraux répétitivement erronés, mais bien d'un conflit d'où la raison sort perdante, privée de son efficace, supplantée par un désir écrasant. Aristote nous

l'enseigne : « quand bien même l'intellect ordonne et que la pensée prescrit de fuir ou de rechercher telle chose, on ne se meut pas pour autant, mais on agit selon l'appétit, comme le font les acratiques » (*De l'âme*, III, 9, 433a 1-3). C'est dire que le principe de réflexion morale est ici, comme cela s'observe dans la psychopathie, préservé, bien qu'il soit ensuite « étourdi ». L'acratique – ou l'incontinent – se retrouve « égaré par l'affection qui tend à le mettre hors de lui-même » (EN, VII, 9, 1151a 15-26), écarté de « la raison droite » (EN, VII, 9, 1151a 15-26), dominé par le poids de cette affection qui le détourne de la voie que lui avait indiquée sa raison à l'issue d'une délibération, qui est bonne, mais dont il n'use pas. Reste à savoir, à ce stade de notre enquête, pourquoi le psychopathe n'use pas de cette délibération. Pourquoi, autrement dit, cède-t-il ainsi par acrasie ? Pour quelle(s) raison(s) son « âme » ne s'en tient-elle pas à sa droite raison ?

D'aucuns diront probablement qu'il convient, pour répondre convenablement, d'aller chercher du côté de la maîtrise de soi (*self-control*), ou d'une faiblesse de la volonté débouchant sur une impulsivité cognitive et comportementale. La psychopathie serait l'indice d'une sorte de *folie de la volonté* réductible au dysfonctionnement des mécanismes volitifs déterminant et régulant l'action. À l'instar du kleptomane qui ne peut se retenir d'assouvir l'impérieux désir qu'il a de voler tout en sachant qu'il ne devrait pas voler, le psychopathe serait tout bonnement incapable de résister à certains de ses désirs les plus pressants. Jouet d'une pulsionnalité qu'il ne maîtrise pas, ou rarement, il finit (parfois) par être terrassé par les désirs que sa raison lui avait pourtant proscrits.

Cette première hypothèse, si elle est séduisante, se heurte toutefois à au moins deux obstacles. D'une part, il est très difficile d'établir une ligne de démarcation nette permettant de distinguer l'*incapacité* à se contrôler du *manque d'effort* mis en œuvre pour se contrôler. Par exemple, nous n'avons pour l'heure aucun moyen de savoir pourquoi un psychopathe qui échoue à un test d'autocontrôle échoue. Est-ce parce qu'il ne peut pas se contrôler, ou bien parce qu'il n'a pas déployé les efforts nécessaires pour se contrôler comme il l'aurait pu ? D'autre part, même lorsque nous sommes capables de trancher pour l'une ou l'autre de ces deux options, notamment en faveur d'un authentique déficit sur le plan des capacités d'inhibition, c'est à chaque fois, comme dans le cas des crises d'épilepsie, sur la base d'une identification d'anomalies cérébrales clairement objectivables. Or, dans la mesure où il n'y a, jusqu'à preuve du contraire, aucune anomalie de ce genre dans la psychopathie, la charge de la preuve se situe du côté des tenants de cette position. Ces derniers doivent nous expliquer en quoi il serait plausible de considérer que c'est parce qu'ils sont incapables d'exercer un contrôle suffisant sur leurs désirs que les psychopathes n'agissent pas en conformité avec les prescriptions de leur intellect. Et même en supposant que nous découvrions un jour suffisamment d'éléments cliniques militant en faveur d'une espèce d'impuissance de la volonté, il resterait à affronter, de nouveau, deux problèmes qui, là encore, sont loin d'être négligeables.

D'abord, on fera remarquer que les conclusions qui s'ensuivraient supposeraient de dissocier purement et simplement la raison de la volonté, comme si la seconde – qui serait alors purement arbitraire – était indépendante de la première. Or, on s'en doute, cela paraît, au moins dans les cas non neurologiques, difficile à soutenir. Un acte voulu n'est-il pas en effet nécessairement un acte guidé par la raison ? Le vouloir n'est-il pas un phénomène qui, d'une façon ou d'une autre, est obligatoirement déterminé *dans* et *par* l'acte de penser ? Ensuite, on attirera l'attention sur les implications que comporte le fait de conclure à une abolition (ou même à une altération) du libre arbitre fondée sur la mise en évidence d'un cerveau défectueux. On notera ainsi, dans le sillage de Joëlle Proust (2005, p. 240), que si des dysfonctionnements neurocognitifs tendent effectivement à déterminer nos comportements (par exemple en nous déposant, de temps à autre, de notre pouvoir de veto), le fonctionnement neurocognitif d'un « cerveau sain » le fera tout aussi sûrement (et implacablement) – de la même façon, d'ailleurs, que si les gènes expliquent les comportements dits « anormaux », ils expliquent aussi les comportements dits « normaux ».

AFFECTIVITÉ ET MORALITÉ DANS LA PSYCHOPATHIE

Face à ces difficultés, nous pourrions être tentés de nous diriger vers une explication que nous taxerons, pour aller vite, de néosentimentaliste. Ses partisans sont les héritiers du philosophe des lumières écossais David Hume (2002 [1751]), pour qui « la moralité est déterminée par le sentiment » (p. 144). Leur thèse est que ce sont nos états affectifs qui conditionnent la teneur de nos jugements de valeur, et en particulier nos jugements moraux. Ils soutiennent à cet égard que les choix acratiques relevés dans la psychopathie résultent d'un déficit affectif, avec des émotions nettement émoussées, voire quasiment annihilées. Or, cela peut vouloir dire plusieurs choses.

Premièrement, que le fait de raisonner totalement de sang-froid, sans le « secours du sentiment », ne met pas en branle l'action. La simple connaissance des principes moraux, mécaniquement investis, ne suffit pas. Il manque « l'étincelle (*spark*) » (Hoffman, 2008) affective qui pousse à la prosocialité. C'est que, comme le disait Hume, la raison est, en elle-même, « inerte ». Elle ne peut rien sans les passions. Ce sont les ressentis émotionnels distinctifs qui constituent les véritables éléments *moteurs* de nos inclinations et de nos agirs moraux. En somme, une compréhension purement cognitive, déconnectée de tout vécu affectif, n'est pas à elle seule motivante, puisque notre conscience morale ne nous interpelle que si nous nous sentons *émotionnellement concernés*. Deuxièmement, et plus fondamentalement encore, il faut comprendre qu'il n'y aurait pas de comportements éthiques si les émotions et les qualités phénoménales qui les accompagnent n'existaient pas. Notre vie affective endosse en réalité un rôle non seulement conatif (une portée motivationnelle), mais aussi, et, avant tout, cognitif. Hume, encore lui, insistait ainsi sur le fait que les émotions nous permettent de tracer les « distinctions morales » qui sous-tendent nos choix. Ce sont précisément, d'après cette conception, nos états affectifs qui conditionnent (au moins partiellement) la valence de nos jugements moraux. Ils fondent litté-

ralement la morale, en édifient les principes. Nos appréciations éthico-morales ne reposent finalement jamais sur des raisonnements froidement rationnels, mais plutôt, si l'on nous permet de citer une nouvelle fois (et une dernière fois) Hume, sur des *sentiments d'approbation* ou de *désapprobation*, de *complaisance* ou de *dégout* (2002 [1751]). Au fond, pourrait-on résumer avec Deonna et Teroni (2008), les valeurs morales sont pour nous ce qu'elles sont « en vertu de leurs liens effectifs avec nos vies affectives » (p. 106), de telle sorte que toute personne affectivement émoussée finira par avoir du mal à *valoriser*, c'est-à-dire à percevoir et à attribuer des valeurs (Tappolet, 2000; de Sousa, 1987, p. 45).

On l'aura compris, l'idée que défendent ces psychologues et philosophes humiens n'est autre que celle consistant à avancer que « c'est le centre indispensable de l'aptitude à valoriser » (Jaworska, 2004, p. 95) qui, dans la psychopathie, serait perturbé. Il manquerait cette forme « d'engagement émotionnel » (Jaworska, 2004, p. 95) qui confère des « poids différents » (Damasio, 1994, p. 81) aux différentes situations perçues, aux diverses possibilités qui s'offrent à nous. Et il s'ensuivrait donc finalement que le psychopathe *sait* qu'il est mal de faire *x* sans être pour autant motivé à agir en conséquence, et donc à modeler son choix à la lumière de ce qu'il sait, puisque ce savoir reste purement abstrait, théorique, seulement indicatif (voire anecdotique), désincarné, non *affecté* par le vécu émotionnel tapissant d'une certaine valeur les contenus idéiques. On pourrait même dire, pour reprendre une nuance souvent rappelée (voir par exemple Blair, Mitchell et Blair, 2005), que si le psychopathe *sait* qu'il est mal de faire *x*, il ne le *sent* pas pour autant, si bien que le problème ne serait pas tant celui d'une *incompétence morale* que celui d'une *indifférence morale*.

Quoi qu'il en soit, on saisit, pour récapituler, que l'acrasie proprement psychopathique – si tant est que l'on puisse effectivement parler d'acrasie – ne résulte pas de la non-résistance à une passion qui déborderait la raison et ses conclusions, mais bien, en un certain sens, d'une absence de « passion », d'une anesthésie émotionnelle qui *permet*, ou qui *autorise* l'élan irrationnel. La passion ne désobéit pas à la raison ; elle ne l'épaule pas assez. Et en des termes plotiniens, nous pourrions même déclarer que ce n'est pas parce que « l'âme » du psychopathe n'est *pas assez impassible* que ce dernier est intempérant, mais c'est au contraire parce que cette dernière est *trop impassible*, insuffisamment affectée.

PSYCHOPATHIE ET RAISON PRATIQUE

Il existe une autre manière de résoudre ce décalage entre ce que le sujet atteint de psychopathie (re)connait et ce qu'il fait. Celle-ci passe par un rejet de la thèse – tacitement admise dans l'approche antirationaliste esquissée à l'instant – voulant que le psychopathe délibère « bien ». Il s'agit plus précisément de remettre en cause l'idée d'une raison qui accèderait suffisamment régulièrement à la conclusion (je ne dois pas faire *x*) qui s'impose. On s'en doute, une telle grille de lecture repose sur l'invocation de dysfonctionnements cognitifs censés influencer sur la rationalité pratique des décisions que peuvent prendre les psychopathes. Elle attire notre attention sur le fait qu'un regard plus circonspect révèle que

sous l'apparente rationalité de ces derniers se cache une rationalité pratique au minimum altérée. Différentes études sont généralement citées pour assoir cette conception. Il est notamment fait mention des liens avérés, systématiquement retrouvés, entre la présence d'une personnalité psychopathique et la présence de « mauvaises décisions », ainsi que de capacités mnésiques et attentionnelles nettement plus faibles (voir par exemple Hiatt et Newman, 2006) qui rendraient notamment difficile, pour les individus concernés, la détection, dans certaines situations, de paramètres susceptibles de les inciter à entreprendre des changements par rapport à l'action qu'ils envisagent (Lykken, 1957; Schmauk, 1990; Jurjako et Malatesti, 2016). Mais ces déficits dits « cognitifs » viendraient surtout contrarier les compétences morales des psychopathes (Kennett, 2002; 2006; voir aussi Maibom, 2005; 2010). L'une des hypothèses phares suggère ainsi que ces derniers ne seraient plus capables, lorsqu'ils se trouvent pour ainsi dire plongés dans le feu de l'action, de maintenir dans leur espace cognitif de travail les informations nécessaires à la production de décisions pertinentes, raisonnables. Leurs facultés attentionnelles amoindries les handicaperaient à vrai dire non seulement dans la conservation et l'intégration effectives des informations pertinentes pour l'action, mais elles perturberaient également le processus de sélection et de hiérarchisation des stimuli permettant la lecture efficiente des situations (que celles-ci soient ou non morales). Il s'ensuit, rapporte-t-on habituellement, qu'ils n'associent pas suffisamment leurs choix à des raisons morales qui ne sont pas *directement* et *explicitement* liées à l'objectif sur lequel ils sont cyclopiquement – et momentanément – fixés (Koenigs et Newman 2013). Aussi serait-il faux de dire qu'ils ont du mal à ajuster leurs comportements à la lumière des résultats négatifs de ces comportements, dans la mesure où ces résultats ne sont pas – ou plus – véritablement appréhendés, perçus, captés sur le plan attentionnel. L'individu sait d'une manière *générale* qu'il est mal de faire x, qu'il ne devrait pas le faire, mais lors de la prise de décision, il n'a finalement plus « en tête » ce principe (qui ne peut donc pas faire office de principe à l'action), de sorte que l'on peut dire qu'il n'a pas conscience, au moment de décider, de toutes les variables à prendre en considération ; il n'a pas à l'esprit la variété des raisons concurrentes susceptibles d'être intégrées au processus délibératif (lequel apparaît donc nécessairement bancal, et débouche sur un mauvais choix).

Pour bien comprendre ce qui se joue ici au niveau cognitif, et en guise d'illustration, intéressons-nous, au risque d'être inutilement didactique, à la structure du syllogisme pratique telle que l'a définie Aristote. Prenons deux prémisses : *il ne faut pas faire de mal à autrui* (majeure) et *je suis en train de faire du mal à autrui* (mineure). L'agent – atteint de psychopathie – doit par conséquent arrêter ce qu'il est *en train de faire à autrui*. Imaginons néanmoins qu'il poursuive malgré tout son action préjudiciable pour autrui, malgré l'interdiction édictée par la (sa) raison ; bien qu'il n'agisse pas *par* ignorance, il agit toutefois *en état* d'ignorance, et son ignorance semble porter sur la conclusion du syllogisme. Pourquoi ? Eh bien, parce que si l'agent ignorait la majeure, alors il ignorerait précisément ce qui constitue l'objet de la science morale. Ce qui nous conduirait tout droit vers l'intellectualisme moral socratique, en ceci qu'il serait, dans ce cas, postulé que le sujet ne sait pas qu'il ne faut pas faire de mal à autrui. Or,

nous l'avons vu, rien, ou quasiment rien, n'indique qu'il y ait, dans la psychopathie, ce genre d'ignorance morale. Cependant, en déduire que notre agent ignore la mineure n'est pas non plus tenable. Aristote, dans son *Éthique à Nicomaque*, nous dit bien que la mineure est en acte, et cela paraît bien nécessaire s'il doit y avoir une action : pour cause, si l'individu ne reconnaît pas qu'il est en train de faire du mal à autrui, il n'agira ni de manière encratique (comme il le ferait en respectant l'interdiction morale), ni de manière acratique (comme il le ferait en cédant à la passion), pour la simple et bonne raison qu'il ne fera rien, puisque rien ne se passe si la mineure n'est pas en place (l'acratique a, de fait, besoin de la mineure, qu'il lie à la majeure stipulant qu'il ne faut pas faire de mal à autrui [EN, 1147a 31-35]). Reste donc, comme nous l'avons dit, qu'il ignore, en un certain sens, la conclusion. Peut-être faudrait-il plutôt dire qu'il n'aboutit pas à la conclusion parce qu'il perd de vue la prémisse (en l'occurrence, ici, morale) qui y conduit. Il peut certes parfaitement énoncer la conclusion, de même que *l'ivrogne* – pour reprendre une figure qu'affectionne Aristote – peut citer les thèses kantienne des *Fondements de la métaphysique des mœurs* en tant qu'il n'est pas dépourvu des facultés cognitives et de raisonnement nécessaires. Malgré tout, il n'en demeure pas moins qu'il n'agit pas en conséquence, étant entendu que dès lors qu'il sort d'un contexte théorique, du cadre de la récitation abstraite ou de l'expérimentation en situation de laboratoire, il ne parvient pas à utiliser à bon escient cette connaissance, à la garder sur le devant de la scène mentale, et donc à s'en servir pour agir adéquatement, conformément à ce que cette connaissance par ailleurs acquise et théoriquement maîtrisée lui suggère.

L'intérêt de cette perspective cognitiviste, on le voit, c'est qu'elle permet, à l'instar de l'approche émotiviste, de préserver l'idée (empiriquement fondée) selon laquelle la psychopathie n'est pas associée à une connaissance morale lacunaire, sans pour autant concéder que cette connaissance morale implique un choix moral solidaire de la connaissance qui le sous-tend. Le fait est que l'on peut donc dire que l'individu avec un diagnostic de psychopathie possède, à un moment donné, un certain discernement (il sait, de manière générale, qu'il est mal de faire *x*). Mais ce discernement *disparaît* au moment de la prise de décision : le sujet ayant, en quelque sorte, dans le feu de l'action et des multiples signaux à traiter, ou de la multitude des variables à considérer, *oublié* qu'il est mal de faire du mal à autrui. Son savoir serait d'une certaine façon *suspendu*.

Aussi intéressante et subtile soit-elle, cette théorie présuppose, tout comme les deux autres examinées juste avant, que le fonctionnement psychopathique laisse place à une activité de délibération entreprise en amont des actions privilégiées. Nous avons vu en effet que, dans le cadre de la thèse faisant du psychopathe un « incontinent », il y a bel et bien délibération, le problème étant que l'action ne suit pas les conclusions auxquelles elle donne lieu. Concernant la thèse accordant un rôle déterminant aux déficits émotionnels venant banaliser la conclusion (pertinente) issue du raisonnement pratique, il y a bien, là aussi, délibération, et ce, quand bien même si le sujet ne s'en tient pas à ce vers quoi sa délibération l'a conduit. Enfin, dans le cas de la thèse faisant des choix immo-

raux du psychopathe le produit de lacunes attentionnelles parasitant l'émergence d'une décision morale appropriée, il y a, là encore, quelque chose qui correspond précisément à une délibération. Celle-ci apparaît certes biaisée, et elle échoue dans la formation d'une opinion morale raisonnée, mais il n'en demeure pas moins que le sujet délibère. Or, et c'est là un point qu'il nous faut à présent traiter, la plupart des descriptions cliniques de la personnalité psychopathique laissent envisager qu'il n'est pas évident d'attribuer aux personnes qui en présentent les principales caractéristiques une telle tendance à la délibération.

PSYCHOPATHIE ET DÉFICIT PRUDENTIEL

Il n'est qu'à se référer à la *Psychopathy Checklist Revised* (PCL-R) (Hare et Neumann, 2006), l'outil diagnostique le plus usité tant dans le champ psychiatrique que dans le domaine médico-légal, à partir duquel nous pouvons extraire deux éléments symptomatologiques centraux susceptibles d'étayer l'idée d'une faible inclination à la délibération pratique, et corrélativement d'un probable déficit prudentiel propre à la psychopathie, tel qu'a pu le proposer Watson (2013). Mentionnons premièrement l'item 13 de cet entretien semi-structuré, dénotant « l'incapacité à planifier à long terme et de façon réaliste », et renvoyant ainsi aux difficultés, éprouvées par l'individu, à formuler et à mettre en œuvre des plans et des objectifs réalistes à long terme. Il est en effet établi que le fonctionnement psychopathique tend à être marqué par un certain présentisme associé à une versatilité importante en matière de planifications et de comportements adoptés. L'anticipation vis-à-vis de l'avenir et des conséquences éventuelles des actions entreprises est rare, voire quasi inexistante (Hare, 2003). L'item qui suit porte quant à lui directement la focale sur le concept d'impulsivité, soulignant le caractère impulsif, non prémédité, irréfléchi et irréflexif des conduites dans lesquelles s'engage le sujet présentant une psychopathie. Celui-ci, nous explique-t-on, fait généralement les choses « selon l'impulsion du moment », parce qu'il « en a envie » ou parce qu'une occasion particulière s'offre à lui. Le recours à la mentalisation, mais aussi à l'analyse coûts-bénéfices, est rare (Hare, 2003, p. 43). Cela a pu faire dire à certains (voir notamment Watson, 2013, ou McIlwain, 2010) que les individus aux traits psychopathiques se distinguent des autres criminels par leur manque de prudence : ils commettent des crimes même lorsque le risque de se faire prendre est particulièrement élevé, privilégiant généralement les gros gains difficilement accessibles (ou risqués) aux petits gains sans risques. Les *patterns* décisionnels et comportementaux dont ils font preuve au fameux jeu de l'Iowa (Bechara *et al.*, 1994; Blair *et al.*, 2001; Mitchell *et al.*, 2002) sont à ce titre instructifs (voir Blair, 2013; Jurjako et Malatesti 2016; Koenigs et Newman 2013; Maibom, 2005). Dans cette tâche dite d'apprentissage instrumental (fréquemment convoquée pour illustrer le manque d'inhibition comportemental et de projection dont font preuve les personnes diagnostiquées avec une psychopathie), les participants doivent déployer un ensemble de stratégies pour gagner autant d'argent que possible, sachant que tous peuvent bénéficier de récompenses sous forme de sommes d'argent, ou au contraire subir des pertes financières, en fonction des jeux de cartes pour lesquels ils optent. Le choix d'une carte rapidement lucrative peut ainsi engendrer des pertes à long

terme, alors qu'une autre carte entrainera plus tard davantage de gains. De façon intéressante, par rapport aux groupes témoins, les participants avec psychopathie adoptent plus de « comportements à risques » lors de la sélection des cartes, subissant par conséquent de lourdes pertes en raison de leur tendance à privilégier une main venant immédiatement les enrichir, en dépit de la valeur marchande moindre, sur le long terme, des cartes qu'ils choisissent.

Nous pourrions alors, face à ce genre de résultat, être tentés de mobiliser la thèse voulant que de tels choix, qui desservent l'individu et finissent par le léser sur le long terme, découlent d'une absence de génération des émotions aversives normalement reliées aux conséquences négatives de son choix (c'est-à-dire la perte d'argent). Il paraît également, *a priori*, justifié de partir du principe que lorsqu'il s'agit de décider de la mise à privilégier, l'attention des sujets avec psychopathie n'est pas (ou plus) dirigée vers les conséquences néfastes associées à telle mise. Une troisième voie d'interprétation consiste néanmoins, comme on l'aura compris, à invoquer l'idée d'une carence significative en matière de préoccupation « prudentielle » (Watson, 2013).

Cleckley et Hare ne cessent de le répéter : les psychopathes sont susceptibles de se retrouver dans des problèmes facilement prévisibles et évitables et sont notoirement inconstants dans leurs plans et projets déclarés. Ce « manque de contrôle » et cette « irresponsabilité » sont d'ailleurs corroborés par nombre de travaux usant d'échelles de mesure autorapportées et de batteries d'évaluation neuropsychologiques démontrant les relations étroites entre psychopathie et niveau élevé d'impulsivité – que l'on parle d'impulsivité cognitive ou comportementale (Fox et Hammond, 2017; Hart et Dempster, 1997; Morgan, Gray et Snowden, 2011; Ray, Poythress, Weir et Rickelm, 2009).

Dépeignant les caractéristiques prudentielles des psychopathes, Watson (2013) souligne à juste titre que ces derniers ont le désir de vivre confortablement, et plus spécifiquement de ne pas être incarcérés. Mais, étrangement, cette préoccupation ne parvient pas à entraver les impulsions criminelles, même minimales (c'est-à-dire même lorsque rien de particulièrement tentant ou d'excitant n'est en jeu) (Hare, 2003), qui mettent en danger leur liberté. Et ces mésaventures prudentielles ne suscitent pas davantage de regrets ou d'autoreproches lorsque c'est autrui qui souffre de leurs actions, de sorte qu'il semblerait que le sujet au fonctionnement psychopathique soit aussi « négligeant » envers lui-même qu'envers les autres. Il n'y aurait donc, chez lui, selon ce point de vue, ni une faiblesse de la volonté raisonnée, ni une indifférence (spécifiquement) morale, ni même une irrationalité pratique, mais tout simplement une espèce d'indifférence généralisée pour ce qui peut advenir (pour lui ou pour autrui).

Soit. Mais comment expliquer une telle indifférence ? Et s'agit-il d'ailleurs réellement, à proprement parler, d'une *indifférence* ? Ne sommes-nous pas plutôt face à ce qui ressemble à une sorte d'*insouciance* ? C'est en tous cas ce qu'a l'air de penser Watson (2013, p. 279), bien qu'il ne nomme pas les choses telles qu'elles. Les psychopathes, selon lui, ne se soucient « pas plus de leur présent que

de leur avenir », et ce, parce qu'ils sont, pour reprendre ses termes, « insuffisamment diachroniques ». Leur incapacité – ou leur grande difficulté – à « prendre soin d'eux-mêmes » proviendrait ainsi plus ou moins directement de difficultés plus profondes à se distancier de leurs désirs et impulsions momentanés pour viser des objectifs à long terme, pour se projeter dans l'avenir, se le représenter, en simuler psychiquement les possibilités, mais aussi pour puiser dans le passé, de façon à réactiver certains vécus et à (re)trouver les *engagements* (*commitments*) – normatifs ou autres – qui façonnent l'identité personnelle et orientent les choix de chacun. Watson paraît ici rejoindre l'idée voulant que les individus caractérisés par un fonctionnement de type psychopathique souffrent d'une altération de leur capacité à voyager mentalement dans le temps (Kennett et Matthews, 2009 ; Levy, 2014 ; Vierra, 2016). Il va même jusqu'à parler, en ce sens, d'un manque « d'orientation normative réflexive », vecteur d'importantes difficultés à produire ce que nous pourrions appeler, avec Kennett (2002, p. 234), une « conception étendue et cohérente de nos propres fins ». Pour autant, et même si certaines personnes rapportant des niveaux élevés de traits psychopathiques disent aussi se sentir piégées (*trapped*), emprisonnées dans un indépassable « présent psychologique » (Fraisie, 1978, p. 205), la question de cette supposée « inefficacité temporelle » signalant un éventuel déficit en matière de « *mental time travel* » (orienté vers l'avenir *et* vers le passé) requiert de plus amples investigations. Par ailleurs, et sans qu'il ne soit nécessaire de recourir à l'existence encore largement incertaine (quoique tout à fait plausible) de ce genre de dysfonctionnement, on peut faire l'hypothèse que les égarements prudents dont font généralement preuve les psychopathes ne sont pas tant – ou prioritairement – la résultante d'une épisodicité permanente de l'expérience consciente que celle d'une absence (préalable) d'inclinations à la réflexivité. Nous entendons par là défendre l'idée que la source de cette insouciance avérée est moins à chercher dans l'affaiblissement de cette dynamique mentale à la fois prospective et rétrospective que dans une dégradation ou un effacement de la dynamique réflexive. D'une certaine façon, cela ferait du psychopathe l'incarnation même du personnage irréfexif (*wanton*) que décrit Frankfurt (1971).

PSYCHOPATHIE, RÉFLEXIVITÉ ET IDENTIFICATION

Le terme irréfexif est ici important. Car il ne s'agit pas, comme le rappelle Neuberg (1991), qui est à l'origine de ce néologisme visant à traduire ce que Frankfurt signifiait par « *wanton* », d'employer le terme *irréléchi*, comme le veut une certaine tradition, ou encore le terme *incontinent*. Le premier tend en effet à reconduire une opposition tranchée, et ici impertinente, entre passion et raison. Or, et comme le spécifie Frankfurt, le *wanton*, à l'instar, manifestement, du psychopathe, n'est pas dépourvu de capacités rationnelles, et peut au contraire parfaitement en faire usage. Mais le *wanton* n'est pas non plus un incontinent (ou un acratique), puisqu'il est, par définition, indifférent à ses désirs et n'a de préférence pour aucun d'eux. Il ne saurait donc agir à l'encontre de son meilleur jugement et faire ainsi preuve de faiblesse de volonté. Rappelons à ce propos, pour clarifier les choses, que « la caractéristique essentielle de l'individu irréfexif est que sa volonté l'indiffère. Ses désirs le poussent à faire certaines choses sans

que l'on puisse dire de lui ni qu'il désire être mû par ces désirs ni qu'il préfère être mû par des désirs différents. » (Frankfurt, 1971, p. 13) Il n'a pas de « volitions et de désirs de second ordre » ; il ne se demande pas, autrement dit, quelle devrait être sa volonté ou son désir, dans la mesure où il ne se questionne pas sur la « désirabilité » de ses désirs spontanés. De sorte que la question « comment dois-je vivre (ou me comporter) ? » ne le traverse pas.

À partir de là, il apparaît effectivement tout à fait légitime de considérer l'individu avec différents traits psychopathiques comme un individu manifestant des difficultés à évaluer les situations (morales ou autres) qui s'offrent à lui, puisque pour évaluer convenablement une situation quelconque, encore faut-il déjà soi-même se demander ce qu'il est convenable de faire dans une telle situation – et donc adopter un point de vue réflexif sur sa propre manière de se positionner dans cette situation. Il semble également, et corolairement, justifié de soutenir, comme cela a pu être proposé, que les difficultés du psychopathe à planifier et à se projeter dans le futur peuvent s'expliquer par cette absence – ou cette quasi-absence – de réflexivité. Il semblerait en effet que pour se préoccuper de l'avenir, pour s'interroger à son sujet, il faille déjà s'intéresser minimalement à ce que l'on veut comme avenir, et donc se préoccuper de ses désirs et de ses volitions. Il en est de même pour ce qui est de la présence d'un déficit prudentiel : celui qui manque de prudence ne se soucie guère des conséquences de ses choix ou de ses actions, tandis que le « prudent » est nécessairement amené à s'interroger sur ce qu'il prévoit de faire – ou sur ce qu'il fait, et donc, en un certain sens, sur ce qu'il veut faire, ou sur ce qu'il désire. Pour délibérer sur le caractère risqué ou non d'une action que j'envisage d'effectuer, il a fallu, en amont, que je m'interroge sur ce que *j'envisage* de faire. Or, si l'on suppose que le psychopathe est un irréfexif au sens où Frankfurt l'entend et le développe, alors ce dernier, dans « son insouciance indifférence à l'égard de la tâche qui consiste à évaluer ses propres désirs » (Frankfurt, 1971, p. 16), n'est aucunement engagé dans ce type d'interrogation réflexive. Déficit prudentiel et déficit réflexif apparaissent donc, sous cet angle, intimement liés. Enfin, et pour aller vers un terrain non encore exploré jusqu'à présent, on peut tout à fait concevoir la psychopathie à travers le prisme de l'irréfexivité tout en argumentant en faveur de l'existence associée – consécutive – d'une identité pratique altérée. Nous avons en effet besoin de désirs concurrents pour former quelque chose s'apparentant à ce que Korsgaard (1996) nomme une *unité pratique de soi*, laquelle découle de la « nécessité d'éliminer les conflits entre les divers motifs » qui nous animent (p. 169). La formation de cette unité pratique requiert que nous nous identifions à certaines de nos volitions, ce qui s'avère indispensable dans la construction d'une certaine forme d'identité volitive. Mais tout cela présuppose de pouvoir comparer ses désirs divergents, de les juger et de peser leur valeur respective, de mettre en perspective leurs effets, de façon à pouvoir décider avec lequel nous nous identifions, à partir duquel nous prévoyons de prendre un engagement décisif en acceptant ce désir comme le nôtre et en nous y pliant. En somme, cela présuppose l'exercice effectif de ce pouvoir d'autoévaluation réflexive qui fait, selon nous, précisément défaut dans la psychopathie. Et c'est pourtant bien cette propension que nous avons de nous rapporter évaluativement à nos désirs de

premier ordre, cette manière dont notre *volonté* est engagée à leur égard, cet intérêt que nous accordons – au sens où nous nous y *intéressons* – à de tels désirs qui fait que nous sommes susceptibles de nous engager dans des délibérations morales portant sur ce qu'il convient de faire. Et c'est justement l'absence (ou, comme nous le disions déjà plus haut, la quasi-absence) de cette activité réflexive impliquée dans la délibération pratique qui fait que le psychopathe irréfexif se laisse vivre au gré de ses envies – aussi immorales, et parfois auto-destructrices, soient-elles – en manifestant une indifférence généralisée à l'égard de sa structure motivationnelle.

DÉFICITS ÉMOTIONNELS ET IRRÉFLEXIVITÉ

Néanmoins, il reste à clarifier ce qui sous-tend cette irréfexivité criminogène que nous présumons. Or, sur cette question, il nous semble que le plus pertinent est de revenir au paradigme sentimentaliste. Si celui-ci semble à première vue devoir être abandonné au profit d'une position en faveur du rôle prépondérant accordé aux aspects volitifs (et plus précisément en faveur de l'absence de volitions de second ordre), un examen approfondi indique cependant qu'elle apparaît conciliable avec un positionnement volitionniste. On peut ainsi arguer, en premier lieu, que la difficile génération d'émotions aversives dissuasives accompagnant, dans la psychopathie, la réponse punitive peut étouffer cet *intérêt* que nous tendons à accorder à nos désirs et volitions de premier ordre. À quoi bon, en effet, *s'intéresser* à ce que nous voulons ou à ce que nous désirons – en adoptant une attitude évaluative réflexive à l'égard de ces volitions ou de ces désirs – lorsque les conséquences négatives de nos actions passées ne suscitent aucun émoi ? N'est-ce pas d'ailleurs en ce sens que nous disons ne pas *éprouver* le besoin – ou la nécessité – de se questionner sur telle ou telle chose ? Par ailleurs, on ajoutera, pour reprendre le fil de nos réflexions sur la question de la rationalité spécifiquement morale du psychopathe, que si les considérations d'ordre éthico-moral ne pèsent pas dans ses décisions finales, ce n'est pas parce qu'elles ne sont pas perçues ou non émotionnellement investies une fois aperçues, ou encore parce qu'elles seraient en dernière instance balayées par un désir soudain irrépessible, mais bien parce qu'elles ne sont pas même recherchées, explorées, ou convoquées, et encore moins examinées, évaluées. Le fait est, en réalité, que lorsqu'elles le sont, la neutralité axiologique qui les caractérise rend tout simplement inutile, ou plutôt vide de sens, leur mise en perspective et en concurrence. La cause en est, doit-on le rappeler, que tout se vaut plus ou moins pour le psychopathe. Les troubles émotionnels, qui, d'après de récentes recherches (Bronchain *et al.*, 2019, 2020; Preszler *et al.*, 2018; Tsang et Salekin, 2019; Verschuere *et al.*, 2018), semblent être au cœur du réseau de symptômes et en organiser les interactions mutuelles, le plongent dans un monde désaffecté, décoloré, dépourvu des « significations remplissantes » (Husserl) susceptibles de permettre la hiérarchisation des raisons (prudentielles et morales) appréhendées. Dans un univers le plus souvent « morne » (Sartre, 2000 [1939], p. 87), structurellement indifférencié, dénué des « atmosphères affectives » (Bégout, 2020) dispersant les « ondes significatives » (Bégout, 2020) octroyant aux situations rencontrées et aux raisons motivant l'agir les *nuances qualitatives* qui leur assi-

gnent une valeur, les sujets atteints de psychopathie perdent de vue le sens, l'intérêt de jauger leurs volitions, de juger leurs désirs à l'aune de principes normatifs et/ou prudents passivement mis sur le même piédestal. À quoi bon, finalement, ébaucher un quelconque mouvement délibératif, « entamer un dialogue de l'âme avec elle-même » (Aubenque, 1963, p. 116), quand toutes les pensées et les « variables » contextuelles se valent, et donc quand rien n'a véritablement de valeur ?

CONCLUSION

Nous sommes partis d'un constat empirique : celui de l'existence d'un fossé, particulièrement marqué dans la psychopathie, et paradoxalement peu discuté, entre jugement moral et choix moral ; un fossé qui, d'une part, tend à remettre en cause l'idée selon laquelle les sujets atteints de psychopathie ne disposent pas des concepts moraux – ainsi que des capacités de raisonnement moral – nécessaires à toute imputabilité morale, et suggère d'autre part de sortir des débats majoritairement polarisés autour des critères strictement épistémiques, au détriment des déterminants volitifs de l'action morale. Il fut donc ensuite question de discuter le bien-fondé de chacun des trois principaux types de réponse traditionnellement privilégiés pour élucider le fossé existant entre ce que le psychopathe sait (ou peut en principe, et dans certaines situations, reconnaître) et ce qu'il décide effectivement de faire. Après avoir examiné tour à tour l'hypothèse acratique, l'hypothèse néosentimentaliste et l'hypothèse cognitive, nous avons indiqué qu'examiner de cette manière l'origine du fossé susmentionné présuppose d'attribuer au psychopathe un usage délibératif de son « intelligence » morale. Et comme ce présupposé s'avère, au regard des données cliniques et de recherches présentées, peu plausible, ou tout du moins difficilement défendable, nous en sommes venus, dans un deuxième temps, à montrer en quel sens il peut être judicieux de penser, comme le fait Watson (2013), la psychopathie à l'aune de déficits prudents qui, en déplaçant la focale sur la manière dont les psychopathes font des choix en général, nous invite à concevoir leurs choix immoraux – qui ne sont en vérité pas vraiment, au sens strict, « immoraux » – comme l'expression d'un manque global de prudence plus que d'une indifférence affective, que d'une irrationalité pratique, ou que d'une faiblesse de la volonté par ailleurs convenablement informée. Il restait cependant à préciser ce qui pouvait expliquer un tel déficit prudentiel ; ce qui pouvait rendre compte de cette insouciance, aussi délétère pour autrui que pour le psychopathe lui-même. Aussi avons-nous soutenu, après avoir évoqué la possibilité, pointée par certains auteurs, d'une altération de la faculté de voyage mental dans le temps, qu'il y a de bonnes raisons d'attribuer au psychopathe une forme de déficit sur le plan de l'autoévaluation réflexive. C'est ici que les travaux de Frankfurt (1971) et sa description de la « créature rationnelle » irréflexive se sont révélés être d'une grande fécondité. Ils nous ont conduit à argumenter que cette apparente irréflexivité jouit d'une portée explicative supérieure par rapport à la quasi-totalité des théories concurrentes qui s'affrontent sur l'échiquier des positions concernant la vie (im)morale du psychopathe. À vrai dire, nous avons essayé de montrer que dans la mesure où chacune de ces théories repose sur le

postulat, souvent implicite, d'une capacité d'autoévaluation réflexive intacte, il devient possible de comprendre plusieurs caractéristiques propres à la psychopathie en convoquant simplement, en tant que facteur pathogène primitif, ce manque supposé de réflexivité. Cela n'enlève toutefois absolument rien à la justesse et à l'intérêt de ces théories. Elles peuvent tout à fait coexister avec une conception du psychopathe comme sujet irréflexif, si tant est que soit admise cette idée selon laquelle c'est précisément, comme nous le soutenons, cette irréflexivité qui constitue la base – ou qui permet justement de rendre globalement compte – des déficits jusque-là convoqués pour éclairer les agissements (im)moraux relevés dans cette affection psychopathique. Un type de déficit nous semblait malgré tout résister à cette conception et échapper au pouvoir explicatif de ce critère réflexif. Il s'agissait, nous l'avons vu, du déficit d'ordre émotionnel, que nous avons en réalité, et pour terminer, convoqué pour comprendre d'où pouvait provenir cette absence (ou cette quasi-absence) d'autoévaluation réflexive.

Et c'est, en l'occurrence, sur cette affectivité sinon évanouie, du moins restreinte, qu'il nous semble opportun de finir pour clôturer notre propos. Car c'est en effet cette question de l'expérience émotionnelle (pathologique) du psychopathe qui nous semble devoir être approfondie lorsqu'il s'agit de rechercher ce qui peut distinguer les psychopathes judiciarisés (*unsuccessful psychopaths*) des psychopathes non judiciarisés (*successful psychopaths*) – distinction sur laquelle nous n'avons pu, faute de place et étant donné la complexité du sujet, nous appesantir. Notons en tous cas que plusieurs auteurs, à l'instar de Varga (2015), qui s'inspire des travaux de Watson (2013), stipulent que la différence, loin d'être rattachée à des dysfonctionnements émotionnels, se logerait avant tout sur le plan des capacités d'autocontrôle réflexif. Les *unsuccessful psychopaths*, à la différence de leurs homologues non judiciarisés (et parfois parfaitement intégrés), se caractériseraient ainsi par une incapacité relative à contrôler l'assouvissement de leurs désirs immédiats (Elliott, 1992; McIlwain, 2010). Et cette spécificité résulterait, toujours selon Varga (2015), qui, là encore, s'inscrit dans le sillage de Watson (2013), d'un manque d'*orientation normative réflexive*. Ce manque entraverait d'une part la consolidation, voire, tout simplement, et plus radicalement encore, la formation d'une identité diachronique suffisamment consistante, et favoriserait d'autre part l'engagement dans les conduites impulsives/irréfléchies propulsant ces « types » de psychopathes en prison. Quant à nous, nous posons plutôt l'hypothèse que toute la différence, ou du moins l'origine commune des différences observées entre les « formes » de psychopathie, se situe, au niveau le plus fondamental, sur un plan purement affectif. Notre hypothèse, qui a le mérite de s'accorder avec les caractérisations aujourd'hui majoritairement dimensionnelles – et non pas catégorielles – de la psychopathie, consiste plus précisément à penser ce mode de fonctionnement et ses diverses expressions comportementales et adaptatives le long d'un continuum de difficultés émotionnelles dont les degrés sont susceptibles de prédire, dans une certaine mesure, les choix (im)moraux de ceux dont les tableaux cliniques sont inclus dans ce spectre psychopathique. C'est dire que la structure dimensionnelle – que l'on peut, si l'on veut, décomposer en « profils » – de la psychopa-

thie se superpose à celle de l'intensité des niveaux d'expérience émotionnelle, et donc, à la profondeur ou l'ampleur de la déficience émotionnelle, dont la graduation peut également permettre la formalisation de paliers ou de classes typologiques plus ou moins associés à certains niveaux de réflexivité et, par conséquent, à certains profils comportementaux (prudentiels et moraux) plus ou moins criminogènes.

BIBLIOGRAPHIE

Aharoni, Eyal, Walter Sinnott-Armstrong et Kent A. Kiehl, « Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction », *Journal of Abnormal Psychology*, vol. 121, no. 2, 2012, p. 484-497.

Andershed, Henrik, « Stability and Change of Psychopathic-Traits. What Do We Know? », in Randall T. Salekin et Donald R. Lynam (dir.), *Handbook of Child and Adolescent Psychopathy*, New York, Guilford Press, 2010, p. 233-250.

Anastassiou-Hadjicharalambous, Xenia et David Warden, « Physiologically-Indexed and Self-Perceived Affective Empathy in Conduct-Disordered Children High and Low on Callous-Unemotional Traits », *Child Psychiatry and Human Development*, vol. 39, no. 4, 2008, p. 503-517.

Aristote, *Éthique à Eudème*, trad. Richard, Bodéüs, Paris, Gallimard, 2014.

———, *Éthique à Nicomaque*, trad. Richard, Bodéüs, Paris, Gallimard, 2014.

———, *De l'âme*, trad. Richard, Bodéüs, Paris, GF-Flammarion, 1993.

Asscher, Jessica J. et al., « The Relationship Between Juvenile Psychopathic Traits, Delinquency and (Violent) Recidivism: A Meta Analysis », *Journal of Child Psychology and Psychiatry*, vol. 52, no. 11, 2011, p. 1134-1143.

Baron-Cohen, Simon, *Zero Degrees of Empathy: A New Theory of Human Cruelty*, London, Penguin, 2012.

Bartels, Daniel M. et David A. Pizarro, « The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Responses to Moral Dilemmas », *Cognition*, vol. 121, 2011, p. 154-161.

Bechara, Antoine, Antonio Damasio, Hanna Damasio, Steven W. Anderson, « Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex », *Cognition*, vol. 50, 1994, p. 7-15.

Bégout, Bruce, *Le concept d'ambiance*, Paris, Seuil, 2020.

Blair, Robert James R., « A Cognitive Developmental Approach to Mortality: Investigating the Psychopath », *Cognition*, vol. 57, 1995, p. 1-29.

———, « Responsiveness to Distress Cues in the Child with Psychopathic Tendencies », *Personality and Individual Differences*, vol. 27, no. 1, 1999, p. 135-145.

———, « A Cognitive Neuroscience Perspective on Child and Adolescent Psychopathy », in Randall T. Salekin et Donald R. Lynam (dir.), *Handbook of Child and Adolescent Psychopathy*, New York, Guilford Press, 2010, p. 156-178.

———, « Reduced Sensitivity to Others' Fearful Expressions in Psychopathic Individuals », *Personality and Individual Differences*, vol. 37, no. 6, 2004, p. 1111-1122.

Blair, Robert James R., E. Colledge et Derek G. Mitchell, « Somatic Markers and Response Reversal: Is There Orbitofrontal Cortex Dysfunction in Boys with Psychopathic Tendencies », *Journal of Abnormal Psychology*, vol. 29, 2001, p. 491-498.

Blair, Robert James R., Derek G. Mitchell et Karina Blair, *The Psychopath: Emotion and the Brain*, Oxford, Blackwell Publishing, 2005.

Borg, Jana S. et Walter P. Sinnott-Armstrong, « Do Psychopaths Make Moral Judgments? », in Kent A. Kiehl et Walter P. Sinnott-Armstrong (dir.), *Handbook on Psychopathy and Law*, Oxford, Oxford University Press, 2013, p. 107-128.

Bronchain, Jonathan, Patrick Raynal et Henri Chabrol, « The Network Structure of Psychopathic Personality Traits in a Non-Institutionalized Sample », *Personality and Individual Differences*, vol. 146, 2019, p. 41-45.

———, « Does Sex Influence the Network Structure of Psychopathy? », *Personality Disorders: Theory, Research, and Treatment*, 2020. Advance online publication. <https://doi.org/10.1037/per0000462>

Cheng, Yawei, An-Yi Hung et Jean Decety, « Dissociation Between Affective Sharing and Emotion Understanding in Juvenile Psychopaths », *Development and Psychopathology*, vol. 24, no. 2, 2012, p. 623-636.

Cima, Maaïke, Franca Tonnaer et Marc D. Hauser, « Psychopaths Know Right from Wrong but Don't Care », *Social Cognitive and Affective Neuroscience*, vol. 5, no. 1, 2010, p. 59-67.

Cleckley, Hervey, *The Mask of Sanity*, 1^{re} éd., St. Louis, C.V. Mosby, 1941.

Cooke, David J. et Christine Michie, « Refining the Construct of Psychopathy: Towards a Hierarchical Model », *Psychological assessment*, vol. 13, no. 2, 2001, p. 171-188.

Damasio, Antonio, *Descartes' Error: Emotion, Rationality and the Human Brain*. New York, Putnam, 1994.

Deonna, Julien et Fabrice Teroni, *Qu'est-ce qu'une émotion ?*, Paris, Vrin, 2008.

de Sousa, Ronald, *The Rationality of Emotion*, Cambridge, MIT Press, 1987.

de Wied, Minet, Anton van Boxtel, Walter Matthys et Wim Meeus, « Verbal, Facial and Autonomic Responses to Empathy-Eliciting Film Clips by Disruptive Male Adolescents with High Versus Low Callous-Unemotional Traits », *Journal of Abnormal Child Psychology*, vol. 40, no. 2, 2012, p. 211-223.

Duff, Antony, « Psychopathy and Moral Understanding », *American Philosophical Quarterly*, vol. 14, no. 3, 1977, p. 189-200.

Elliott, Carl, « Diagnosing Blame: Responsibility and the Psychopath », *The Journal of Medicine and Philosophy*, vol. 17, no. 2, 1992, p. 199-214.

Ellis Havelock, *The Criminal*, New York, Scribner & Welford, 1890.

Eslinger, Paul J. et Antonio R. Damasio, « Severe Disturbance of Higher Cognition after Bilateral Frontal Lobe Ablation: Patient EVR », *Neurology*, vol. 35, 1985, p. 1731-1741.

Fox, Siobhan et Sean Hammond, « Investigating the Multivariate Relationship Between Impulsivity and Psychopathy Using Canonical Correlation Analysis », *Personality and Individual Differences*, vol. 111, 2017, p. 187-192.

Fraisse, Paul, « Time and Rhythm Perception », in Edward Carterette and Morton Friedman (dir.), *Perceptual Coding*, New York, Academic Press, 1978, p. 203-254.

Frankfurt, Harry G., « Freedom of the Will and the Concept of a Person », *Journal of Philosophy*, vol. 68, no. 1, 1971, p. 5-20.

Frick, Paul J. et Monica A. Marsee, « Psychopathy and Developmental Pathways to Antisocial Behavior in Youth », in Christopher J. Patrick (dir.), *Handbook of Psychopathy*, New York, Guilford Press, 2018, p. 456-475.

Furnham, Adrian, Yasmine Daoud et Viren Swami, « How to Spot a Psychopath », *Social Psychiatry and Psychiatric Epidemiology*, vol. 44, no. 6, 2009, p. 464-472.

Glenn, Andrea L., Spassena Koleva, Ravi Iyer, Jesse Graham et Peter H. Ditto, « Moral Identity in Psychopathy », *Judgment and Decision Making*, vol. 5, no. 7, 2010, p. 497-505.

Glenn, Andrea L., Adrian Raine et Robert A. Schug, « The Neural Correlates of Moral Decision-Making in Psychopathy », *Molecular Psychiatry*, vol. 14, 2009, p. 5-6.

Greenspan, Patricia, « Responsible Psychopaths », *Philosophical Psychology*, vol. 16, no. 3, 2003, p. 417-429.

Hare, Robert D., *Hare Psychopathy Checklist-Revised (PCL-R)*, 2^e éd., Toronto, Multi-Health Systems, 2003.

Hare, Robert D. et Craig S. Neumann, « The PCL-R Assessment of Psychopathy: Development, Structural Properties, and New Directions », in Christopher J. Patrick (dir.), *Handbook of Psychopathy*, New York, Guilford Press, 2006, p. 58-90.

———, « Psychopathy as a Clinical and Empirical Construct », *Annual Review of Clinical Psychology*, vol. 4, 2008, p. 217-246.

Harris, Grant T., Tracey A. Skilling et Marnie E. Rice, « The Construct of Psychopathy », *Crime and Justice*, vol. 28, 2001, p. 197-264.

Hart, Stephen D. et Rebecca J. Dempster, « Impulsivity and Psychopathy », in Christopher D. Webster et Margaret A. Jackson (dir.), *Impulsivity: Theory, Assessment, and Treatment*, New York, Guilford Press, 1997, p. 212-232.

Hiatt, Kristina et Joseph Newman, « Understanding Psychopathy: The Cognitive Side », in Christopher J. Patrick (dir.), *Handbook of Psychopathy*, New York, Guilford Press, 2006, p. 334-352.

Hodsoll, Sara, Nillie Lavie et Essi Viding, « Emotional Attentional Capture in Children with Conduct Problems: The Role of Callous-Unemotional Traits », *Frontiers in Human Neuroscience*, vol. 8, 2014, p. 570.

Hoffman, Martin, *Empathie et développement moral : les émotions et la justice*, Grenoble, PUG, 2008.

Hume, David, *Enquête sur les principes de la morale*, trad. fr. Michel Malherbe, Paris, Vrin, 2002 (1751).

Husserl, Edmund, *Recherches logiques*, Tome II, 2^e partie, V, § 8, trad. H. Elie, A. L. Kelkel et R. Schérer, Paris, PUF, 2011.

Jaworska, Agnieszka, « Ethical Dilemmas in Neurodegenerative Disease », in Judy Illes (dir.), *Neuroethics: Defining the Issues in Theory, Practice and Policy*, Oxford, Oxford Scholarship Online, 2004, p. 95.

Jurjako, Marco et Luca Malatesti, « Instrumental Rationality in Psychopathy: Implications from Learning Tasks », *Philosophical Psychology*, vol. 29, no. 5, 2016, p. 717-731.

Kennett, Jeanette, « Autism, Empathy and Moral agency », *The Philosophical Quarterly*, vol. 52, no. 208, 2002, p. 340-357.

———, « Do Psychopaths Really Threaten Moral Rationalism? », *Philosophical Explorations*, vol. 9, no. 1, 2006, p. 69-82.

———, « Reasons, Emotion and the Psychopath », in Luca Malatesti et John McMillan (dir.), *Responsibility and Psychopathy: Interfacing Law, Psychiatry and Philosophy*, Oxford: Oxford University Press, 2010, p. 243-259.

Kennett, Jeanette et Steve Matthews, « Mental Time Travel, Agency and Responsibility », in Matthew Broome et Lisa Bortolotti (dir.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, Oxford, Oxford University Press, 2009, p. 327-350.

Kiehl, Kent A. et Walter P. Sinnott-Armstrong, (dir.), *Handbook on Psychopathy and Law*, Oxford, Oxford University Press, 2013.

Koenigs, Michael, Michael David Kruepke, Joshua D. Zeier et Joseph P. Newman, « Utilitarian Moral Judgment in Psychopathy », *Social Cognitive and Affective Neuroscience*, vol. 7, no. 6, 2012, p. 708-714.

Koenigs, Michael et Joseph P. Newman, « The Decision-Making Impairment in Psychopathy: Psychological and Neurobiological Mechanisms », in Kent A. Kiehl et Walter P. Sinnott-Armstrong (dir.), *Handbook on Psychopathy and Law*, Oxford, Oxford University Press, 2013, p. 93-106

Korsgaard, Christine M., *The Sources of Normativity*, Cambridge, Cambridge University Press, 1996.

Kurzban, Robert, Peter DeScioli et Daniel Fein, « Hamilton vs. Kant: Pitting Adaptations for Altruism Against Adaptations for Moral Judgment », *Evolution and Human Behavior*, vol. 33, no. 4, 2012, p. 323-333.

Leistico, Anne-Marie R., Randall T. Salekin, Jamie DeCoster et Richard Rogers, « A Large-Scale Meta-Analysis Relating the Hare Measures of Psychopathy to Antisocial Conduct », *Law and Human Behavior*, vol. 32, no. 1, 2008, p. 28-45.

Levy, Neil, « The Responsibility of the Psychopath Revisited », *Philosophy, Psychiatry, & Psychology*, vol. 14, no. 2, 2008, p. 129-138.

———, « Psychopaths and Blame: The Argument from Content », *Philosophical Psychology*, vol. 27, no. 3, 2014, p. 351-367.

Lykken, David T., « A Study of Anxiety in the Sociopathic Personality », *Journal of Abnormal and Social Psychology*, vol. 55, 1957, p. 6-10.

Malatesti, Luca et John McMillan (dir.), *Responsibility and Psychopathy: Interfacing Law, Psychiatry and Philosophy*, Oxford, Oxford University Press, 2010.

Maibom, Heidi L., « Moral Unreason: The Case of Psychopathy », *Mind & Language*, vol. 20, no. 2, 2005, p. 237-257.

———, « Rationalism, Emotivism and the Psychopath », in Luca Malatesti et John McMillan (dir.), *Responsibility and Psychopathy: Interfacing Law, Psychiatry and Philosophy*, Oxford, Oxford University Press, 2010, p. 305-324.

Marshall, Julia, Ashley L. Watts, Erika J. Frankel et Scott O. Lilienfeld, « Do Psychopathic Individuals Possess a Misaligned Moral Compass? A Meta-Analytic Examination of Psychopathy's Relations with Moral Judgment », *Personality Disorders: Theory, Research, and Treatment*, vol. 9, no. 1, 2018, p. 40.

Marsh, Abigail A. et Robert James R. Blair, « Deficits in Facial Affect Recognition Among Antisocial Populations: A Meta-Analysis », *Neuroscience Biobehavioral Review*, vol. 32, no. 3, 2008, p. 454-65.

Marsh, Abigail A. *et al.*, « Empathic Responsiveness in Amygdala and Anterior Cingulate Cortex in Youths with Psychopathic Traits », *Journal of Child Psychology and Psychiatry*, vol. 54, no. 8, 2013, p. 900-910.

Maudsley, Henry, *Responsibility in Mental Disease*, London, H.S. King, 1873.

McGeer, Victoria., « Varieties of Moral Agency: Lessons from Autism and Psychopathy », in Walter Sinnott-Armstrong (dir.), *Moral Psychology, Volume 3: The Neuroscience of Morality*, Cambridge, MIT Press, 2008, p. 227-257.

McIlwain, Doris, « Living Strangely in Time: Emotions, Masks and Moral in Psychopathically-Inclined People », *European Journal of Analytic Philosophy*, vol. 6, no. 1, 2010, p. 7594.

Morgan, Joanne E., Nicolas S. Gray et Robert J. Snowden, « The Relationship Between Psychopathy and Impulsivity: A Multi-Impulsivity Measurement Approach », *Personality and Individual Differences*, vol. 51, no. 4, 2011, p. 429-434.

Morse, Stephen J., « Psychopathy and Criminal Responsibility », *Neuroethics*, vol. 1, no. 3, 2008, p. 205-212.

Murphy, Jeffrie, « Moral Death: A Kantian Essay on Psychopathy », *Ethics*, vol. 82, 1972, p. 284-298.

Neuberg, Marc (dir.), *Théorie de l'action. Textes majeurs de la philosophie analytique de l'action*, Bruxelles : Mardaga, 1991.

Nichols, Shaun, « How Psychopaths Threaten Moral Rationalism, or is It Irrational to be Amoral? », *The Monist*, vol. 85, no. 2, 2002, p. 285-303.

Patrick, Christopher J. (dir.), *Handbook of Psychopathy*, New York, Guilford Press, 2018.

Petrican, Raluca et Chris Burris, « The Infernal Now: Linking Temporal Inefficacy to Cognitive Ability and Social Adjustment », *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, vol. 43, no. 1, 2011, p. 52-62.

Preszler, Jonathan *et al.* « Network Analysis of Psychopathy in Forensic Patients », *Journal of Abnormal Psychology*, vol. 127, no. 2, 2018, p. 171-182.

Prichard, James C., *A Treatise on Insanity*, London, Sherwood, Gilbert and Piper, 1835.

Prinz, Jesse, *The Emotional Construction of Morals*, Oxford, Oxford University Press, 2007.

Proust, Joëlle, *La Nature de la volonté*, Paris, Gallimard, 2005.

Ray, James V., Norman G. Poythress, John M. Weir et Angela Rickelm, « Relationships Between Psychopathy and Impulsivity in the Domain of Self-Reported Personality Features », *Personality and Individual Differences*, vol. 46, no. 2, 2009, p. 83-87.

Schmauk, Frank J., « Punishment, Arousal, and Avoidance Learning in Psychopaths », *Journal of Abnormal Child Psychology*, vol. 18, 1990, p. 451-463.

Schramme, Thomas (dir.), *Being Amoral: Psychopathy and Moral Incapacity*, Cambridge, MIT Press, 2014.

Syngelaki, Eva *et al.*, « Affective Startle Potentiation in Juvenile Offenders: The Role of Conduct Problems and Psychopathic Traits », *Social Neuroscience*, vol. 8, no. 2, 2013, p. 112-121.

Talbert, Matthew, « Blame and Responsiveness to Moral Reasons: Are Psychopaths Blame-worthy? », *Pacific Philosophical Quarterly*, vol. 89, no. 4, 2008, p. 516-535.

Tappolet, Christine, *Émotions et valeurs*, Paris, PUF, 2000.

Tassy, Sebastien, Christine Deruelle, Julien Mancini, Samuel Leistedt et Bruno Wicker, « High Levels of Psychopathic Traits Alter Moral Choice but Not Moral Judgment », *Frontiers in Human Neuroscience*, vol. 7, 2013a, p. 229.

Tassy, Sebastien, Olivier Oullier, Julien Mancini et Bruno Wicker, « Discrepancies Between Judgment and Choice of Action in Moral Dilemmas », *Frontiers in Psychology*, vol. 4, 2013b, p. 250.

Tassy, Sebastien, Olivier Oullier, Michel Cermolacce et Bruno Wicker, « Do Psychopathic Patients Use Their DLPFC When Making Decisions in Moral Dilemmas? », *Molecular Psychiatry*, vol. 14, 2009, p. 908-909.

Tsang, Siny, Randall T. Salekin, « The Network of Psychopathic Personality Traits: A Network Analysis of Four Self-Report Measures of Psychopathy », *Personality Disorders: Theory, Research, and Treatment*, vol. 10, no. 3, 2019, p. 246-256.

Varga, Somogy, « Identifications, Volitions and the Case of Successful Psychopaths », *dialectica*, vol. 69, no. 1, 2015, p. 87-106.

Verschuere, Bruno *et al.* « What Features of Psychopathy Might be Central? A Network Analysis of the Psychopathy Checklist-Revised (PCL-R) in Three Large Samples », *Journal of Abnormal Psychology*, vol. 127, no. 1, 2018, p. 51-65.

Vierra, Andrew, «Psychopathy, Mental Time Travel, and Legal Responsibility», *Neuroethics*, vol. 9, no. 2, 2016, p. 129-136.

Watson, Gary, «Psychopathic Agency and Prudential Deficits », *Proceedings of the Aristotelian Society*, vol. 113, no. 3, 2013, p. 269-292.

IS DANIEL A MONSTER? REFLECTIONS ON DANIEL A. BELL AND WANG PEI'S "SUBORDINATION WITHOUT CRUELTY" THESIS

RAINER EBERT

FELLOW, OXFORD CENTRE FOR ANIMAL ETHICS, AND INTERNATIONAL RESEARCH ASSOCIATE, UNIVERSITY OF DAR ES SALAAM

VALÉRY GIROUX

ASSOCIATE DIRECTOR, CENTRE DE RECHERCHE EN ÉTHIQUE AND ADJUNCT PROFESSOR, LAW FACULTY, UNIVERSITÉ DE MONTRÉAL

ANGIE PEPPER

LECTURER, SCHOOL OF HUMANITIES AND SOCIAL SCIENCES, RESEARCH CENTRE FOR PRACTICAL PHILOSOPHY, THEOLOGY AND RELIGION, UNIVERSITY OF ROEHAMPTON

KRISTIN VOIGT

ASSOCIATE PROFESSOR, INSTITUTE FOR HEALTH AND SOCIAL POLICY & DEPARTMENT OF PHILOSOPHY, MCGILL UNIVERSITY

ABSTRACT:

Daniel Bell and Wang Pei's recent monograph, *Just Hierarchy*, seeks to defend hierarchical relationships against more egalitarian alternatives. This paper addresses their argument, offered in one chapter of the book, in favour of a hierarchical relationship between human and nonhuman animals. This relationship, Bell and Pei argue, should conform to what they call "subordination without cruelty:" it is permissible to subordinate and exploit animals for human ends, provided that we do not treat them cruelly. We focus on three aspects of their view: their argument for a hierarchical view; their understanding of cruelty; and their account of the heightened duties they claim we owe to nonhuman animals who are intelligent, domesticated, and/or "cute." We argue that the reasons that Bell and Pei offer fail to support their conclusions, and that, even if one accepts a hierarchical view, the conclusions that Bell and Pei draw about the permissibility of practices such as killing animals for food do not follow. We conclude by emphasizing philosophers' responsibility to thoroughly test their arguments and to engage with existing debates, especially when the practices they seek to justify involve harms of great magnitude.

RÉSUMÉ :

Dans leur récent ouvrage, *Just Hierarchy*, Daniel Bell et Wang Pei se donnent pour tâche de défendre le principe de hiérarchie au détriment des relations plus égalitaires entre les êtres. Dans cet article, nous commentons la position qu'ils défendent dans un chapitre de leur livre consacré à la relation qu'entretiennent les êtres humains avec les autres animaux. Cette relation, soutiennent-ils, devrait relever de ce qu'ils appellent une « subordination non cruelle ». Il serait selon eux légitime d'assujettir et d'exploiter des animaux pour des fins humaines, à condition que nous ne les traitons pas cruellement. Dans cet article nous nous concentrons sur trois aspects de leur position : leur défense du principe de hiérarchie, leur manière d'envisager la notion de cruauté et ce qu'ils regardent comme

les devoirs accrus que nous aurions, jugent-ils, à l'égard des animaux intelligents, domestiqués et/ou « mignons ». Selon nous, les arguments avancés par Bell et Pei ne permettent pas de défendre adéquatement leur thèse ; en outre, même en admettant leur hypothèse sur la hiérarchie, les conséquences qu'ils en tirent au sujet de la légitimité morale de certaines activités telles que l'élevage pour la viande ne nous semblent pas recevables. Nous concluons en soulignant la responsabilité qui incombe aux philosophes de mettre scrupuleusement à l'épreuve leurs arguments et de connaître les débats actuels dans le domaine qu'ils abordent, tout particulièrement lorsque les pratiques qu'ils entreprennent de justifier sont la cause d'immenses souffrances.

1. INTRODUCTION

Daniel Bell and Wang Pei's recent monograph, *Just Hierarchy*, sets out to defend hierarchical relationships against more egalitarian alternatives. This paper is concerned with their argument in favour of a hierarchical relationship between human and nonhuman animals (ch. 4). This relationship, Bell and Pei argue, should conform to what they call "subordination without cruelty," which means that it is permissible to subordinate and exploit animals for human ends, provided that we do not treat them cruelly. Bell and Pei argue that subordination without cruelty remedies the implausibility of theories of animal rights that are strongly egalitarian, such as Donaldson and Kymlicka's (2011) account, which starts from the assumption that all sentient animals have basic inviolable rights. To Bell and Pei, this view and other egalitarian accounts like it are untenable because they yield counterintuitive implications. Most pertinently, they suggest that such egalitarian, rights-based views would brand Bell a monster because of his past interactions with cats, which involve, for instance, eating them, failing to prevent a fatal accident, and physically punishing them for unwelcome behaviour (p. 159).¹ This, they believe, shows that egalitarian animal-rights theory is fundamentally flawed: Daniel clearly is not a monster, which means that "the problem lies with the theory" (p. 159).

We want to resist this move. Specifically, we want to resist the claim that if Daniel is not a monster, then egalitarian animal-rights theory is wrong. It is true that Daniel's actions constitute socially accepted behaviour, and many readers will probably agree that Daniel's actions do not make him a monster: animals are widely seen as the kind of beings that it is permissible to kill, eat, harm, or exploit for human benefit.² It is also true that a theory such as Donaldson and Kymlicka's rules out the norms and the actions Bell and Pei describe. Yet it does not follow that Daniel is a *monster*. It does, however, follow that Daniel's actions were *morally wrong*.

Bell and Pei's argument might reasonably be interpreted as an attempt to justify predominant social attitudes about the status of animals and to legitimate much of the status quo. Our critical response to Bell and Pei's argument proceeds from the insight that if we are to make progress in our thinking about interspecies justice, we must be open to the possibility that many, perhaps even most, of the ways in which we currently interact with nonhuman animals are morally wrong. There is now a substantial and growing philosophical literature that seeks to establish that point and, more positively, to describe what just relations between human and nonhuman animals might look like. Throughout this paper, we will argue that by not questioning the norms that currently inform our behaviour vis-à-vis nonhuman animals and by not engaging with the substantial literature that challenges their claims, Bell and Pei fail to defend the position they assert in this chapter. As we argue at the end of this paper, philosophers have a responsibility to do better, especially when their arguments relate to practices that involve harms of such significant magnitude.

Our critical response to Bell and Pei begins by challenging their argument in favour of a hierarchical view (section 2). Our aim is to suggest that egalitarian animal-rights views do not necessarily have the radical implications they suggest, and to argue that their uncritical reliance on intuition is deeply problematic. We then turn our attention to their understanding of cruelty. We argue that their definition of cruelty is underspecified and undefended and that it potentially permits all manner of practices that many people regard as “cruel” (section 3). We then critically discuss their claim that we have special duties towards domesticated, intelligent, and cute animals (section 4). Lastly, we situate Bell and Pei’s argument in the context of contemporary animal ethics and suggest that they fail to show due diligence in approaching a topic that has serious implications for some of the most vulnerable beings (section 5).

2. BELL AND PEI’S ARGUMENT IN FAVOUR OF A HIERARCHICAL ACCOUNT

Instead of offering a positive argument in support of their hierarchical account, Bell and Pei suggest that the egalitarian alternative, which grants equal rights to nonhuman animals, has counterintuitive implications that make it untenable. First, if human and nonhuman animals had equal rights, we should feel the same level of moral outrage when their rights are violated. But clearly, they say, this is not how we in fact respond. Second, if human and nonhuman animals had the same basic rights, then animal-rights advocates should want to mete out the same punishment to perpetrators of rights violations against nonhuman animals as to perpetrators of right violations against human animals. Referencing Donaldson and Kymlicka (2011), they argue that even animal-rights advocates do not accept this implication. From this, they conclude that “it definitely doesn’t make sense to say that animals and humans have equal rights” (p. 151).

A number of points are worth making here. First, even though Donaldson and Kymlicka’s *Zoopolis* is one of Bell and Pei’s main reference points throughout this chapter, they misconstrue Donaldson and Kymlicka’s position. Donaldson and Kymlicka are not in principle opposed to punishing violations of nonhuman animals’ rights just as harshly as those of humans’ rights. They say explicitly that they look forward to a future where the killing of any animal, human or nonhuman, is treated as murder (Donaldson and Kymlicka, 2011, p. 133). They also suggest that, under current conditions, there could even be a case for punishing the killing of nonhuman animals *more* harshly than the killing of humans, because harsher punishments could provide much-needed deterrence in the absence of established norms that prohibit such killings (Donaldson and Kymlicka, 2011, p. 133).

Second, recognizing that an individual has moral rights does not imply any specific conclusions about how violations of those rights ought to be legally punished. For one, granting equal moral status to individuals does not necessarily imply granting them all the exact same moral rights. Indeed, the specific rights granted to protect the interests of individuals may vary according to the

interests those individuals have.³ Also, penal sanctions reflect a range of concerns and objectives that may have different implications depending on the victim's species. For example, if punishment should reflect judgments about individual blameworthiness, we may have to take into account the degree to which the transgressor flouted prevailing norms: this would suggest that violating the rights of nonhuman animals should be subject to less severe punishment than violations of human rights (Donaldson and Kymlicka, 2011, p. 133).⁴ Another example is reparation: the human victim of a rights violation could be consoled by a harsh sentence for the perpetrator while a harsh sentence brings no such relief to a nonhuman victim.

More broadly, Bell and Pei's rejection of egalitarian views, which paves the way for their hierarchical account, reflects the uncritical reliance on intuitions that characterizes their methodology. Bell and Pei are quick to assert a claim, state that it is obvious or intuitive, and then move on as if the claim had been defended. This methodology is deeply flawed: our intuitions about how we ought to treat (different kinds of) nonhuman animals are shaped by currently prevailing norms, which unreflectively sanction the use and exploitation of animals. Given that what is at stake in this debate is precisely whether or not these norms must be rejected, we cannot simply take intuitions as given.

3. ON BELL AND PEI'S UNDERSTANDING OF CRUELTY

The broad position defended by Bell and Pei is that although animals are the moral subordinates of humans, this does not give us a licence to treat animals cruelly. We have already seen why Bell and Pei reject the claim that animals have equal negative rights. In this section, we want to consider what counts as cruelty on their view and why they think cruelty towards animals is morally prohibited. We will argue that their view is woefully underdeveloped, and that their understanding of cruelty is underpinned by the undefended assumption that it is permissible to make animals suffer if it suits human ends.

After surveying some religious traditions that affirm a moral hierarchy of species with humans at the top—Christianity, Confucianism, and Buddhism—Bell and Pei argue that none of these traditions permits humans to do as they please with animals (p. 152–53). Indeed, according to each of these worldviews, humans must extend some compassion to other animals and refrain from abusing them. Bell and Pei's central motivation for invoking these traditions is to illustrate that a commitment to subordination is not a commitment to cruelty. This prepares the ground for their own view, which is that while animals do not have the same rights as humans, we nonetheless have “a duty to avoid cruelty to animals, both because animals can suffer and because those who are cruel to animals are more likely to be cruel to humans” (p. 153). In short, Bell and Pei endorse both a direct duty argument against cruelty (we owe it to the animals not to make them suffer) and an indirect duty argument against cruelty (we owe to ourselves and other humans not to make animals suffer).

So far, so good. The idea that it is wrong to be cruel to animals is one that will be accepted by almost all. The trouble comes when we try to specify precisely what counts as cruel. Bell and Pei offer no concrete definition of cruelty, nor do they give us much by way of examples, except to point out that there is a lot of cruelty in intensive animal farming. In general, their view seems to be “You’ll know cruelty when you see it.” However, this lack of precision is deeply problematic when taken in conjunction with what they deem to be compassionate (or at least not cruel) treatment. Let us look at some examples.

Many animal advocates argue that it is “immeasurably cruel” (Korsgaard, 2018, p. 228) to induce illness in animals and test noxious substances on them for the purpose of medical research. Yet Bell and Pei suggest that this is morally acceptable (although we should try to use animals who have less in common with humans, such as rats⁵) (p. 154). Similarly, many would argue that “whacking” animals to discipline them, which is what Daniel does to the family cat Didi (p. 158), is also cruel. Yet, while Daniel feels “somewhat guilty” about the incident, he notes that he succeeded in getting Didi to refrain from acting out again (p. 158). Bell and Pei also describe how Didi was in “terrible pain for the last couple of weeks of his life and...remained immobile on the staircase, without any appetite” (p. 158). In this case, one wonders why Didi was not provided with adequate pain relief or granted the benefit of a good death, and one might plausibly argue that allowing an animal to suffer needlessly for two weeks cannot be anything but cruel. Lastly, and perhaps most disturbingly, Bell and Pei describe how Xiao Bao (a cat not much liked by the family) leapt to his death from an unsecured, twenty-second-floor window after being sprayed with water—a punishment for scratching the furniture (p. 157). While Bell and Pei find this incident regrettable, they do not think Xiao Bao was treated cruelly. Since no one intended for Xiao Bao to be harmed, we might agree with Bell and Pei, but that would depend on us accepting the controversial thesis that cruelty is always intentional and never a result of negligence.

At this point, Bell and Pei could (though note that they themselves do not) respond by making a familiar argumentative move: cruelty is bad because it involves *unnecessary* suffering, but some suffering can be justified when it is necessary to achieve important or reasonable human ends. On this view, while necessary suffering is regrettable, it is not cruel, and it is permissible because the costs to humans of refraining from imposing suffering on animals are deemed to be too high. So, while torturing animals for fun will be ruled out as cruel, making animals suffer in the name of science may be considered necessary when the cost to human health of abstaining from experiments on animals would be too great.⁶

Importantly, most of the suffering that Bell and Pei seem to endorse cannot be justified by the appeal to the idea that it is necessary. Was it necessary for Daniel to whack Didi as punishment for biting him? Most compassionate animal guardians will argue that it was not. There are many ways to discipline animals (and children), and one need not resort to physical violence. What human end

was served by not securing adequate pain relief or euthanasia for Didi? It presumably saved Daniel and his family money, but unless they were destitute this could hardly serve as a justification. Similarly, it was not necessary to spray Xiao Bao with water. This act cost Xiao Bao his life. What was the cost to Daniel and his family of not spraying Xiao Bao? Scratched furniture. Could they have tried to change Xiao Bao's behaviour without terrifying him? Certainly. Since all of these instances of making animals suffer are arguably unnecessary, it is difficult to see why they are not acts of cruelty.

Ultimately, on Bell and Pei's account, cruelty is determined not by how much suffering an animal experiences, but rather by whether suffering is necessary for what *they* judge to be reasonable human ends. In defence of Bell and Pei, one might argue that they explicitly reject the idea that animals are *mere* means for human ends, when, for example, they argue that animals "are ends in themselves who have their own goals and can suffer, and we have a duty to minimize their suffering" (p. 165). But it is difficult to square this thought with their view that it is permissible to eat animals, use them for medical research, physically punish them, confine them, leave them to die painfully, terrify them, or cull them (p. 176). What makes these actions permissible, on Bell and Pei's view, is precisely that we are permitted to treat animals as mere means to human ends. Few, if any, of these harms are strictly speaking *necessary*—we need not farm animals, eat animals, keep them as companions, or use them for medical research. They occur precisely because animals are regarded as mere means to our ends.

In sum, while Bell and Pei pay lip service to the familiar idea that cruelty is immoral, their judgment that Daniel's behaviour is permissible, as well as their endorsement of practices that inevitably involve harm to animals, indicates that their understanding of cruelty does little to protect animals against the worst of humanity. Bell and Pei are right to note that attributing a subordinate moral status to animals does not give us a licence to treat them cruelly, but their own view seems to assume that subordination does in fact allow us to treat animals as we please and that we are permitted to cause animal suffering even when it is not necessary to satisfy *important* human ends.

4. HEIGHTENED DUTIES IN RELATION TO INTELLIGENT, DOMESTICATED, AND CUTE ANIMALS

For Bell and Pei, human beings are morally superior to members of other species. This, they think, entitles human beings to use nonhuman animals for various purposes (as long as cruelty is avoided). However, while all nonhuman animals are subordinate to humans, Bell and Pei suggest that there is a hierarchy among nonhuman animals. The authors argue that we have more—and more stringent—duties to those animals who meet one (or more) of these three conditions: they have certain "human" capacities (in particular, intelligence); they are under our care rather than independent; they are "cute." Some of these are direct duties; others are indirect duties and are ultimately grounded in a concern for human interests. Let us take a closer look at how Bell and Pei justify these duties.

Bell and Pei argue that it is “intuitively obvious” (p. 154) that we owe more to animals with “human-like traits such as intelligence, empathy, self-consciousness, and the ability to be aware of themselves as distinct entities with a past and a future” (p. 154). Other things being equal, animals endowed with sophisticated cognitive capacities would therefore be placed higher on the scale of moral value than others. For example, when it comes to using animals to test medical treatments that can cure human illness, “obviously we should choose a rat over an ape” (p. 154–5).

The question of what role, if any, cognitive capacities should play in determining individuals’ moral status has been examined in the literature on animal ethics, in the debate on basic equality in the human context, and in disability studies scholarship, none of which Bell and Pei draw on.⁷ Many animal ethicists⁸ argue that species membership is just as morally arbitrary as many of the other categories we use to draw distinctions between individuals, such as gender or race. On the face of it, cognitive capacities might seem a more promising way to justify differences in moral status between members of different species. However, this position clearly has troubling implications: if intelligence or other superior cognitive capacities ground higher moral status among nonhuman animals, it seems that this would also commit us to drawing similar distinctions among *human* beings who differ in relation to these capacities (see Dombrowski, 1997; Pluhar, 1995). In light of all the work that has been done on this question in animal ethics and in disability studies, we cannot simply take for granted that less intelligent beings have less moral value or that their like interests need not be considered equally.

The second consideration that Bell and Pei identify when it comes to the relative status of different nonhuman animals is relational: they argue that we have positive duties towards the animals we keep and for whom we are responsible (although they reject Donaldson and Kymlicka’s argument that these duties would include making them our fellow citizens). Bell and Pei, it seems, see these special duties as both direct and indirect: they are grounded in the fact that “they depend on us for their well-being,” but they are also based on the fact that “our relations with them enhance human well-being” (p. 159).

In response, note that Bell and Pei’s discussion equivocates between the wider category of domesticated animals and the narrower one of companion animals: the duties they consider are owed to our “pets,” such as cats and dogs, and not, say, turkeys—the latter, Bell and Pei suggest, can continue to be served at Thanksgiving dinners (p. 173). In fact, Bell and Pei consider *only* companion animals when discussing our duties towards animals in our care. This is no coincidence, since their primary concern is not animals or their interests, but rather the way in which our treatment of animals affects human well-being (p. 159). Our special duties “to” companion animals are really duties toward our fellow humans. Indeed, Bell and Pei take up the Kantian argument that cruelty to animals should be avoided because people who behave violently toward animals tend to behave violently toward humans as well.

A similar, indirect duty argument is at play when Bell and Pei argue that we have greater obligations to the animals we find cute (cuteness is typically the reason we welcome some domesticated animals into our homes but eat others) because expressing compassion for them has the potential to increase social harmony and to encourage us to be kinder towards members of the most vulnerable human groups (p. 161). For Bell and Pei, the reason that we owe less to, at one end of the spectrum, ugly and harmful insects than to the family dog, at the other end of the spectrum, is not only that the different types of relations we have with each entail different duties, but also that the mistreatment of the former is not as detrimental to humans as the mistreatment of the latter (p. 144).

Note that Bell and Pei's reference to insects is misleading because animal-rights theorists focus on *sentient* animals; to what extent insects meet this requirement is a contested question, so it is unclear what role they can play in the argument (Elwood, 2011; DeGrazia, 2020). Bell and Pei's rhetoric also seems to suggest that "ugliness" and harmfulness go hand in hand (p. 144, 175), which is certainly not always the case. Furthermore, whether indirect benefits for human beings can indeed be obtained from protecting certain animals, and whether and to what extent these benefits are greater when the animals in question are "cute," depend on empirical claims that, as Bell and Pei themselves acknowledge, are not settled. Most importantly, they offer no support for the idea that animals' moral status and their fundamental rights should depend on the value—including the aesthetic value—that others place on them. We would clearly not accept this claim in the human context without further argument. The burden of proof is on Bell and Pei to demonstrate why such criteria should be relevant when it comes to nonhuman animals.

5. BACK TO THE BROADER PICTURE: ALTERNATIVE ACCOUNTS AND PROFESSIONAL ETHICS

For the longest time, philosophers have paid very little attention to the moral status of nonhuman animals. If they mentioned animals at all, they did so mostly just to contrast them with humans and to make the point that only humans are appropriate objects of ultimate moral concern. Rarely did philosophers extend serious ethical consideration to other animals, and when they did so, they were at best perceived as eccentric or sentimental.

In recent decades, this situation has changed drastically, especially since the publication of the seminal books by Peter Singer and Tom Regan, respectively, *Animal Liberation*, in 1975, and *The Case for Animal Rights*, in 1983. Today, nonhuman animals and their moral status are the subject of a lively academic debate. There is a continuous stream of conferences and workshops dedicated, sometimes exclusively, to the moral status of nonhuman animals, and there are various journals that specialize in publishing work on the human-animal relationship. The number of books and articles that deal with issues in animal ethics has exponentially increased:

The most comprehensive bibliography of writings on the moral status of animals lists only ninety-four works in the first 1970 years of the Christian era, and 240 works between 1970 and 1988, when the bibliography was completed. The tally now would probably be in the thousands (Singer, 2003).

Similarly, political theorist Robert Garner observes that “more has been written on animal ethics in the past three decades or so than in the previous 2000 years” (Garner, 2005, p. 4).⁹

Within this debate, egalitarianism, in various forms, occupies a central position and has been argued for with great philosophical care and in great detail. Bell and Pei focus on the influential theory of animal rights defended by Donaldson and Kymlicka, just to dismiss it rather flippantly. Other key theories, such as Regan’s, are not even mentioned. Bell and Pei suggest that defenders of the idea of moral equality between humans and other animals, such as Donaldson and Kymlicka, “write as though they feel the same sense of outrage when the basic rights of animals are violated” (p. 150) as when the basic rights of humans are violated, implying that Donaldson and Kymlicka really do *not*. Bell and Pei do not stop at effectively accusing their philosophical opponents of being disingenuous. They further suggest that Donaldson and Kymlicka have not thought through their position. Had Donaldson and Kymlicka “pushed their argument to its logical conclusion” (p. 151), they would have to “call for life-long imprisonment of animal meat eaters or people who kill insects” (p. 150). Regardless of the validity of this inference, over which we have already expressed our doubts, the suggestion that egalitarians do not fully understand or appreciate the implications of their respective views is rather presumptuous, arguably offensive, and certainly uncharitable. It is safe to assume that Donaldson and Kymlicka—as well as Regan and others for that matter—are genuine about their feelings and have a solid grasp of what the egalitarian position implies, and that they are prepared to accept those implications.

To be sure, Bell and Pei are not alone in their rejection of egalitarianism when it comes to nonhuman animals. In thinking that the egalitarian position has implications that are “absurd,” they are in good company, and they are right to draw attention, in particular, to life-or-death situations:

Should electric cars be programmed to swerve away from animals even if it endangers the life of the driver? Should scientists be forced to end all animal experimentation even if...almost all the medicine we use has been tested on animals? (p. 151)

The possibility that the egalitarian response to such questions might prevent us from valuing human over nonhuman life has troubled philosophers, some enough to be convinced that egalitarianism has no plausible answers. Among them are Jeff McMahan and Shelly Kagan who, like Bell and Pei, have consequently rejected egalitarianism in favour of hierarchy.¹⁰

In *The Ethics of Killing* (2002), McMahan writes that “Radical Egalitarianism must imply that, if other things are equal, the killing of a fish is as seriously wrong as the killing of a person—which is absurd” (p. 208). He then goes on to propose a hierarchical moral theory that has two tiers, an egalitarian tier for persons and a consequentialist one for nonpersons (a more sophisticated version of Robert Nozick’s “utilitarianism for animals, Kantianism for people”). Kagan instead proposes a gradual account of moral status. His motivation, though, is largely the same. In his recent book, *How to Count Animals, More or Less* (2019), which rightly receives much attention, he argues that egalitarians cannot avoid “the unacceptable conclusion that it is immoral for Tom to kill fish, rabbits, or deer to keep himself alive” (p. 184). This, according to Kagan, is a strong reason to reject the view that all animals with moral status have the very same moral status.

Quite unlike Bell and Pei, however, both McMahan and Kagan are very much in agreement with egalitarians in condemning the way in which we currently treat nonhuman animals, specifically when we use them for food. For example, McMahan argues that

if one is to be justified in killing an animal, one must, *at a minimum*, have a purpose that is sufficiently serious to outweigh the animal’s time-relative interest in continuing to live. While I accept that there are certain purposes that can meet this burden of justification—for example, certain medical experiments—there are other widespread human practices, such as the social practice of eating meat, that I believe cannot (McMahan, 2002, p. 203).

Similarly, Kagan writes:

Our treatment of animals is a moral horror of unspeakable proportions, staggering the imagination. Absolutely nothing that I say here is intended to offer any sort of justification for the myriad appalling and utterly unacceptable ways in which we mistreat, abuse, and torture animals (Kagan, 2019, p. 5).

This, he makes clear, also includes meat eating, which is “tremendously...difficult to justify—typically, perhaps, impossible” (Kagan, 2019, p. 228).

Reasonable people may agree that “if it’s a question of life and death, humans should have priority” (p. 154), but it is a long way from there to justifying eating cats and turkeys for social or cultural reasons (p. 168-173). It is telling that philosophers other than Bell and Pei who hold hierarchical views, such as McMahan and Kagan, have not made that leap—and not at all surprising. The extensive animal-ethics literature contains a wealth of arguments against the practice of eating meat and the industry it sustains. At this point, anybody seeking to defend the production and consumption of meat has a heavy burden to meet. The immense harm to sentient beings that is involved makes acknowl-

edging that burden a moral imperative, rather than just an academic requirement. It should no longer be acceptable for professional philosophers to nonchalantly give their stamp of approval to the practice of meat eating (even if done rarely and only on socially or culturally significant occasions) without *substantially* engaging with the existing arguments against that practice. To suggest to the reader, who may be unsuspecting and unfamiliar with the relevant literature, that it is enough to note that “some people really do love meat” (p. 172) and that meat eating promotes a sense of community (p. 173) is not just a distortion of where we are in animal ethics, but irresponsible.

In fact, there may be a general lesson for philosophers here. When writing as experts, philosophers, like any other experts, have a certain responsibility and should be acutely aware of the potential harm their words may do. The more harm that may come from a particular claim and its communication to a wide audience, the higher the burden of proof, and the more important it is for philosophers to make sure, *in good faith*, that they have given due and charitable consideration to relevant arguments in the literature.

NOTES

- ¹ Page numbers that are not preceded by additional bibliographical information refer to Bell & Pei (2020).
- ² There are of course cross-cultural differences, and different cultures have different norms when it comes to the treatment of *specific* animals, such as cats or cows.
- ³ What is more, moving from moral to legal rights requires a complex deliberative process that takes into account different democratic considerations, such as security, public health, public order, feasibility, and the effect of the proposed law on other important rights and liberties.
- ⁴ See also Fricker's (2007, p. 107) discussion of the role moral knowledge plays in judgments about blameworthiness.
- ⁵ It should go without saying, but sadly it needs to be said, that rats are intelligent sentient beings with the capacity to feel pain and to suffer. It is difficult to see why they might suffer or be harmed less by medical research than primates would. From the rat's point of view, being poisoned, operated on, and made ill are terrible things to have happen to them. And though rats may differ from gorillas, rats have just as much of an interest in not having these harms imposed on them.
- ⁶ We do not accept the claim that making nonhuman animals suffer for medical research is a case of necessary suffering. This example serves only to illustrate how the logic of this line of argument is supposed to work.
- ⁷ For discussion in the animal-ethics literature, see DeGrazia (1996); Ebert (2018); Kagan (2019); Korsgaard (2018); Rachels (1990); and Singer (1975). In the basic-equality literature, see Arneson (1999, 2015). For discussion in disability theory, see Carlson (2020); Crary (2020); Kittay (2005); Silvers (2012); and Wasserman et al. (2017).
- ⁸ See, for example, Cavalieri (2001, chapter 4); Francione (2000, chapter 4); Horta (2018); Jaquet (2022); McMahan (2002, chapter 3); and Singer (1975, chapter 2; 2009; 2011, chapter 3).
- ⁹ This growing debate notwithstanding, some philosophers still simply assume that animals have a lesser moral status than humans (e.g., Waldron, 2017), and some still use comparisons to animals as *reductiones ad absurdum* (e.g., Arneson, 2015).
- ¹⁰ This is, if one understands egalitarianism to be the view that all individuals with any moral standing at all have the very same moral status. In another sense, McMahan is an egalitarian. His egalitarianism, however, is restricted to persons.

REFERENCES

Arneson, Richard J., "What, If Anything, Renders All Humans Morally Equal?", in Dale Jamieson (ed.), *Peter Singer and His Critics*, Oxford, Blackwell, 1999, p. 103-129.

———, "Basic Equality: Neither Acceptable nor Rejectable", in Uwe Steinhoff (ed.), *Do All Persons Have Equal Moral Worth? On "Basic Equality" and Equal Respect and Concern*, New York, Oxford University Press, 2015, p. 30-52.

Bell, Daniel A. and Wang Pei, *Just Hierarchy: Why Social Hierarchies Matter in China and the Rest of the World*, Princeton, Princeton University Press, 2020.

Carlson, Licia, "On Moral Status and Intellectual Disability: Challenging and Expanding the Debates", in Adam Cureton and David Wasserman (eds.), *The Oxford Handbook of Philosophy and Disability*, Oxford, Oxford University Press, 2020, p. 482-497.

Cavalieri, Paola, *The Animal Question: Why Nonhuman Animals Deserve Human Rights*, Oxford, Oxford University Press, 2001.

Crary, Alice, "Cognitive Disability and Moral Status", in Adam Cureton and David Wasserman (eds.), *The Oxford Handbook of Philosophy and Disability*, Oxford, Oxford University Press, 2020, p. 451-466.

DeGrazia, David, *Taking Animals Seriously: Mental Life and Moral Status*, Cambridge, Cambridge University Press, 1996.

———, "Sentience and Consciousness as Bases for Attributing Interests and Moral Status: Considering the Evidence and Speculating Slightly Beyond", in L. Syd M. Johnson, Andrew Fenton and Adam Shriver (eds.), *Neuroethics and Nonhuman Animals*, Cham, Springer International Publishing, 2020, p. 17-31.

Dombrowski, Daniel A., *Babies and Beasts: The Argument from Marginal Cases*, Urbana, University of Illinois Press, 1997.

Donaldson, Sue, and Will Kymlicka, *Zoopolis: A Political Theory of Animal Rights*, New York, Oxford University Press, 2011.

Ebert, Rainer, "Mental-Threshold Egalitarianism: How Not to Ground Full Moral Status", *Social Theory and Practice*, vol. 44, no. 1, 2018, p. 75-93.

Elwood, Robert W., "Pain and Suffering in Invertebrates?", *ILAR*, vol. 52, no. 2, 2011, p. 175-184.

Francione, Gary, *Introduction to Animal Rights: Your Child or the Dog?*, Philadelphia, Temple University Press, 2000.

Fricker, Miranda, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford, Oxford University Press, 2007.

Garner, Robert, *Animal Ethics*, Cambridge, Polity, 2005.

Horta, Oscar, "Moral Considerability and the Argument from Relevance", *Journal of Agricultural and Environmental Ethics*, vol. 31, no. 3, 2018, p. 369-388.

Jaquet, François, “What’s Wrong with Speciesism”, *The Journal of Value Inquiry*, vol. 56, no. 3, 2022, p. 395-408.

Kagan, Shelly, *How to Count Animals, More or Less*, New York, Oxford University Press, 2019.

Kittay, Eva, “At the Margins of Moral Personhood”, *Ethics*, vol. 116, no. 1, 2005, p. 100-131.

Korsgaard, Christine, *Fellow Creatures: Our Obligations to the Other Animals*, New York, Oxford University Press, 2018.

McMahan, Jeff, *The Ethics of Killing*, New York, Oxford University Press, 2002.

Pluhar, Evelyn B., *Beyond Prejudice: The Moral Significance of Human and Nonhuman Animals*, Durham, Duke University Press, 1995.

Rachels, James, *Created from Animals: The Moral Implications of Darwinism*, New York, Oxford University Press, 1990.

Regan, Tom, *The Case for Animal Rights*, Berkeley, University of California Press, 1983.

Silvers, Anita, “Moral Status: What a Bad Idea!”, *Journal of Intellectual Disability Research*, vol. 56, no. 11, 2012, p. 1014-1025.

Singer, Peter, *Animal Liberation*, New York, New York Review/Random House, 1975.

———, “Animal Liberation at 30”, *The New York Review of Books*, vol. 1, no. 8, (May 15, 2003), <http://www.nybooks.com/articles/archives/2003/may/15/animal-liberation-at-30/> [accessed on March 24, 2021].

———, “Speciesism and Moral Status”, *Metaphilosophy*, vol. 40, no. 3-4, 2009, p. 567-581.

———, *Practical Ethics*, third edition, Cambridge, Cambridge University Press, 2011.

Taylor, Sunaura, *Beasts of Burden: Animal and Disability Liberation*, New York, The New Press, 2017.

Waldron, Jeremy, *One Another’s Equals: The Basis of Human Equality*, Cambridge, MA, The Belknap Press, 2017.

Wasserman, David, Adrienne Asch, Jeffrey Blustein and Daniel Putman, “Cognitive Disability and Moral Status”, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2017.

JUSTICE, IMPARTIALITY, AND EQUALITY IN THE ALLOCATION OF SCARCE VACCINES: A REPLY TO SAUNDERS

HUGH V. MCLACHLAN

GLASGOW CALEDONIAN UNIVERSITY

ABSTRACT:

This paper is a reply to Saunders's criticism of my previously published nonconsequentialist policy proposal regarding the use of a lottery for the distribution of scarce vaccine by the state in the face of an influenza pandemic. I argued there that, on the grounds of justice, the state should distribute some of the scarce vaccine it might hold to some of its healthcare employees and the rest to citizens randomly and equally on the principle of a lottery. Central to Saunders's criticism is the claim that I mistakenly failed to take into account morally relevant differences in need and productive capacity between potential recipients of the vaccine. Central to my response here is that whether or not need and productive capacity are morally relevant factors depends on who or what is distributing the vaccine, to whom they are distributing it, and why they are doing so. For instance, discrimination between people in the distribution of publicly owned vaccine that is distributed as a prophylactic on the basis of their age is unjust whether or not it might be morally justifiable on other grounds.

RÉSUMÉ :

Cet article est une réponse à la critique de Saunders de ma proposition de politique non conséquentialiste publiée précédemment concernant l'utilisation d'une loterie pour la distribution de vaccins rares par l'État face à une pandémie de grippe. J'y ai soutenu que, pour des raisons de justice, l'État devrait distribuer une partie du vaccin rare qu'il pourrait détenir à certains de ses employés de la santé et le reste aux citoyens de manière aléatoire et égale sur le principe d'une loterie. Au cœur de la critique de Saunders se trouve l'affirmation selon laquelle j'ai omis à tort de prendre en compte les différences moralement pertinentes dans les besoins et la capacité de production entre les receveurs potentiels du vaccin. Le point central de ma réponse ici est que le fait que le besoin et la capacité de production soient ou non des facteurs moralement pertinents dépend de qui ou de quoi distribue le vaccin, à qui ils le distribuent et pourquoi ils le font. Par exemple, la discrimination entre les personnes dans la distribution de vaccins publics distribués à titre prophylactique sur la base de leur âge est injuste, qu'elle soit ou non moralement justifiable pour d'autres motifs.

INTRODUCTION

In a recent article in this journal, Saunders has criticized my proposed nonconsequentialist policy regarding the use of a lottery for the distribution of scarce vaccine by the state in the face of an influenza pandemic (Saunders, 2018). I argued that, in some circumstances, on the grounds of justice, the state should distribute the scarce vaccine it holds to citizens randomly and equally on the principle of a lottery (McLachlan, 2012, 2015). Saunders writes:

The proposal to give “equal chances” to everyone is not as simple or as appealing as it first appears. Any policy will have to be sensitive to differences, in need and productive capacity, between people... Thus, while I believe that policymakers should show equal concern for all citizens, I suggest that this may permit, or even require, departures from equality in the distribution of (chances of getting) vaccine. (Saunders, 2018, p. 80)

As we shall see, Saunders’s claim about the need to be sensitive to differences in need and productive capacity is not as simple or as appealing as it might first appear.

I shall answer the specific questions that he addresses to me and thereby restate, amplify, and develop my position. Saunders’s criticism contains some very strong points about the ethical allocation of scarce vaccine in general. Recognition and accommodation of them serves to strengthen and develop rather than discredit my initial policy proposal.

A NONCONSEQUENTIALIST ETHICAL JUSTIFICATION OF THE DISTRIBUTION OF SCARCE VACCINE IN A PANDEMIC

My argument was, at root, a response to the tacit consequentialist rationale of the official UK governmental policy on vaccinations in influenza pandemics as expressed in the UK Contingency Pandemic Plan of 2005. It stated: “The priority in an influenza pandemic is to reduce the impact on public health (i.e. reduce illness and save lives). Interventions will therefore be applied where they will achieve maximum health benefit.”¹ The recommended policy involved dividing the population into priority groups including groups of high medical risk and age groups, as well as certain healthcare workers and providers of essential services. I argued that, in certain assumed conditions, if there were an influenza pandemic or an impending one and the state held a scarce supply of vaccine, it should, in fulfilment of its duty of care towards its employees, distribute it firstly to the front-line care providers who might be at risk of catching influenza in their work at the state’s behest of providing healthcare. Thereafter,

if there is not sufficient vaccine to give all other citizens equally an effective dose, the state should give them all an equal chance of receiving an effective dose.... This would be the just thing to do because the

state has a duty to treat each and all of its citizens impartially and they have a corresponding right to such impartial treatment. (McLachlan, 2012, p. 318)

In my view, it matters ethically not only what consequences ensue from our actions, but also how and why we do what we do (Emanuel and Wertheimer, 2006; Jansen and Wall, 2021; Williams and Dawson, 2020; Colgrove, 2019). It matters, too, who does what is done and in what roles they perform their actions. If particular envisaged circumstances—such as, for instance, the outcome of the effects of the taking of doses of vaccine by particular groups of people—happen to be particularly morally good, it does not follow that, by a sort of imaginative reverse engineering, we can deduce that all or any of us must have a moral duty to bring or to try to bring these circumstances about. What distribution of vaccine would produce the morally best outcome? This is a different question from, How ought we to distribute it? This latter question will have different correct answers depending upon whether “we” are, for instance, rich philanthropic private citizens or, for instance, officials or agents of the state.

When people perform particular roles, they are morally constrained as well as morally authorized by rights and duties that are intrinsic to these roles and by the nature of these (McLachlan and Swales, 1999). They do not have the moral freedom of action that they have as normal private citizens. For instance, it is not the business of the members of a jury to decide what, overall, would be the morally best outcome of declaring particular defendants to be innocent or guilty when they produce their verdicts. It is not the business of lecturers to allow a consideration of the moral desirability or undesirability of the overall consequences of passing or failing particular students to influence the marks they choose to give to the particular essays of these students that they are responsible for marking. Similarly, those who perform roles in the government and other such agencies of the state are constrained as well as legitimated by their occupancy of such roles. The primary end of their action is not the morally best outcome, and even when what they do does happen to produce the morally best outcome, that in itself is no automatic vindication of their policy making.

For instance, incumbents of such roles have a moral duty to act justly. Members of governments, like other public officials, agents, and other members of agencies, have a moral duty to treat citizens impartially as individuals in the distribution of benefits and burdens that are individually experienced. Citizens have a corresponding moral right to be treated impartially by them. Each citizen should be treated the same unless there is a morally relevant difference pertaining to them that justifies the different treatment of them. This is the nature of distributive justice.

Justice demands impartial treatment of individual people as individual people. It pertains to differences in their treatment that are related to and proportionate to morally relevant differences in their personal features, conditions, and circumstances. This is different from discriminating among particular people on

the basis of average differences among features and factors that pertain to the different categories or groups that the particular individuals belong to. Hence, I consider the priority given to age categories in the distribution of vaccine that is proposed in the UK Contingency plan for influenza pandemics and that is practiced currently in the UK in the distribution of vaccine against COVID-19 to be unjust (Savulescu and Cameron, 2020; Hughes, 2020; Peterson, 2008).

Consider, for instance, the role of lecturers with regard to the treatment of their students in, say, the marking of their essays. As a matter of justice, they are required to treat them the same and to give them the same mark unless there are morally relevant reasons for treating them differently, in which case discrimination is not only justifiable, but morally obligatory. Acting justly does not merely mean treating one's students the same unless there is a moral justification of some sort for treating them differently. There must be a morally relevant justification. That means treating them differently only when the difference in treatment is morally justified by a corresponding and proportionate difference pertaining to the merits of the essays that are submitted by them. Notice that the allocation of different treatment, of, say, benefits and burdens, is just or unjust regardless of the consequences of that particular allocation. The virtue of distributive justice does not lie in the consequences of any particular distribution of burdens and benefits, but in its appropriateness. With regard to distributive justice, one might say that the proof of the pudding is in the recipe rather than in the eating.

Among the assumptions I made were that anyone might die from the influenza and that all might want to live. The chance of a vaccine that might prevent them dying from the influenza was of value to each and every one of them. That people who are, for instance, older or who belong to a particular sex or racial or ethnic origin might be, on average, more prone to infection or more likely to die if infected is an irrelevance. The presumptions that the state should make are that all people need to live in order to do what they want to do and to enjoy what they want to enjoy, and that they all want to live. A dose of the scarce vaccine might prevent influenza from killing any particular one of them, as far as any one of them can know. They each have valid claims to receive one, between which it is not proper for the state to discriminate.

If we knew in advance what particular people would die or suffer from serious illness if they were to catch, say, influenza or COVID-19, there might be a good case for saying that, on the grounds of justice, they, regardless of their age, should be given priority in the allocation of an appropriate vaccine. However, we do not have and cannot have that information. Most of those who have died from COVID-19 so far or who have suffered serious illness have been old, but, with regard to justice, this is irrelevant. Not all old people who caught it have died or suffered serious illness. Not all who have died or suffered from serious illness have been old. Moreover, the pattern of the pandemic might change from what it has been in the past. The virus might change. Available effective remedial treatment might change. Any of us might catch COVID-19. If we do, for all that

we know, we might die or have serious long-term or short-term consequences to our health or our life expectancy. Whether or not any of us needs a vaccine, we might all want to have one and equally have a legitimate interest in getting one or, at least, the chance of getting one. Younger people who catch COVID-19 might show slight symptoms now, but, for all we know, might die in years to come much younger than they would have done had they not caught it. They might suffer later in life from appalling illnesses and disabilities. We do not have the sort of knowledge available to us that would morally justify discrimination between the younger and the older in the distribution of vaccine.

To give priority in a pandemic in the allocation of vaccine against influenza or COVID-19 on the basis of membership of age groups is unjust. It might be expedient to allocate vaccines in such a way. It might even, in some circumstances, have consequences that morally permit or even morally compel the adoption of such a policy, but it would remain, nonetheless, an unjust one. It is like giving awards to particular school children on the basis of the school they attended or the socioeconomic class of their parents. If one age group has, per capita, a higher rate of deaths or serious illness among those who catch COVID-19 or influenza than a younger age group, it does not mean that all those who are in each group share the same chances of dying or of becoming seriously ill if they catch the disease as their fellow age-group members. Within each age group, there might well be much variation in the chances of the particular members dying or suffering from a serious illness if they catch the disease. Some or many people in the younger group might well have a far greater chance of dying or suffering from a serious illness if they catch it than some or many people in the older group.

What counts as impartiality in the distribution of publicly financed and provided goods or services such as, say, medical surgery, higher education, or vaccines and what counts as morally relevant reasons for discrimination are contentious issues. It depends partly on what we consider to be the proper function of the goods or services in question, to whom the benefit should be bestowed, and for what purpose (Wardrope, 2012). For instance, the need of the recipient is sometimes but not always thought to be a morally relevant reason for discrimination in, say, the distribution of medical surgery, but not in, say, the distribution of places in higher education. Some people—very many, perhaps—think that places in universities should be allocated to those people who can benefit society most from taking them up and that higher education is not primarily for the benefit of students but for the expansion of the economy. Most people, I am sure, would say that medical treatment should be allocated primarily for the benefit of the recipients. We would typically give, say, a heart transplant to someone who needed a new heart and who was certain or very likely to die or be seriously ill without such treatment rather than to someone who did not. However, we would hardly consider it to be just to give heart transplants to, say, everyone in a particular age group before giving one to any to those people in a younger group on the grounds that, in the past, a higher proportion of older people than younger ones needed a heart transplant or, to say a different thing, that a higher

proportion of those who needed a heart transplant in the past were in the older age group. Similarly, we do not and would not give heart transplants to all smokers before giving one to any nonsmoker on the grounds that, on average, smokers are more likely than nonsmokers to need one. To act justly, we should give, for instance, heart transplants to those particular people who need one rather than to those who do not, whatever age they are and whether or not they are nonsmokers. We should allocate, when need is the relevant factor, on the basis of the needs of individual potential recipients regardless of the social or natural categories to which they belong.

When vaccines are allocated primarily as a prophylactic, to people who are not in ill, in order to try to ensure that they do not become ill, primarily for their benefit, the notion of “need” is not readily applicable. Healthy people do not need a vaccine against, say, COVID-19, in the way in which some people who currently suffer from a particular illness need, for instance, a heart transplant and those who do not suffer from it do not. However, anyone might catch and die from or suffer severely in the long or short term from catching COVID-19. All of us might want a vaccination as soon as possible against catching it. We each have a legitimate interest in receiving the vaccine. Any request we might explicitly or implicitly put forward deserves consideration. Not all such requests can be immediately acceded to but the grounds for their rejection or the deferral of compliance to them should be providable by any government that wants to act justly and to be seen to do so.

With some vaccines, in some sorts of pandemics, we might all have a very strong interest in the provision of vaccines to each and every person. My argument was, like the official policy to which I was responding, based on the assumption that the vaccine against influenza would be distributed as a prophylactic, for the benefit of its recipients. Where the benefits of particular treatments and services are not intended to be enjoyed solely or primarily by those who receive them, the question of whether and how they can be distributed impartially is more complex (Wardrope, 2012; McLachlan, 2015). Some vaccines can protect those who receive them from becoming ill from catching a particular viral infection without preventing them from becoming asymptomatically infected and infectious. Some vaccines can prevent the recipient from becoming infected and being infectious. Thus, we might benefit significantly from the vaccination of other people if and when their lack of infectiousness can reduce our chances of becoming ill or dying from catching a particular infection.

We might all equally have a very strong interest in the vaccination of all people with whom we might directly or indirectly socially interact. To try to give everyone, citizen and noncitizen alike, who is living within the geographical boundaries of the state an efficacious dose of vaccine as promptly as possible—regardless of the ordering of the distribution to particular people or to particular categories of them—might in some circumstances be a very good policy. It might be a policy that could be considered sufficiently impartial in the circumstances. It might be the morally optimal vaccination policy for a government to

adopt at least with some sorts of vaccines in some or other stages of some sorts of pandemics. It might form part of an optimal package of policies that a government could try to implement.

The state is morally obliged to refrain from wantonly killing each and every one of us but is not similarly morally obliged to do all that is required to try to postpone the deaths of each and every one of us (McLachlan, 2010). To let us die is, normally, morally permissible for the state whether or not there is a pandemic. Whether or not there is a pandemic, particular deaths can be delayed, and particular lives lengthened, very often, only at the expense or risk of hastening other deaths and shortening other lives. Hence, the notion that, in a pandemic or in any other times, the government has a paramount duty to save lives or, indeed, that it has any sort of paramount duty is not plausible. To protect the National Health Service (NHS), for instance, is not inherently ethically more important than, say, to protect the education system or, say, the systems of civil or criminal justice. The postponement of particular deaths does not matter more than or less than, say, the education of children or the livelihoods of adults or the freedom of adults to, for instance, hug, kiss, and enjoy other forms of intimate contact with other consenting adults. Such things matter ethically but they matter differently. They are different morally good things. There is a host of things that are morally permissible for governments to do, a host of things that are morally desirable for them to do, a host of things that are obligatory for them to do, and a host of things that are obligatory for them to refrain from doing. Hence, it is appropriate for governments to try to produce an optimal blend of policies rather than to try to maximize the specified output of any particular prioritized policy or policies, whether or not they are in a pandemic.

Notice that, in some instances, unjust actions and policies might be morally justifiable on grounds other than justice.² For instance, governments might be confronted by a severe emergency to which they do not have, say, the time or imagination to devise a policy that is both a just and an effective solution. Perhaps, in some such cases, there is none. However, they should acknowledge the injustice of such actions and policies. The state, its agents, and its agencies should not lightly violate the moral right to justice, even when the benefits that are unjustly denied or the burdens that are unjustly bestowed are comparatively slight. We, as citizens, should often be prepared to waive our moral right to just treatment in particular circumstances. We should, however, beware that governments do not get into the habit of overriding it.

IMPARTIAL TREATMENT AND DIFFERENT LEVELS OF NEED

(A) Equal Chances and Unequal Need

Saunders writes: “McLachlan argues that the state has a special duty of care towards public healthcare providers, who undertake risks at its behest, which justifies giving them priority. But, with this exception, he suggests that other citizens should be given equal chances of receiving an effective dose of

vaccine....However, it is not obvious that impartiality always requires equal chances, particularly when individuals are differently situated” (Saunders, 2018, p. 67).

To illustrate his point, he asks us to

consider a small-scale example that may illustrate this point. Suppose that Alpha is very sick and needs 100 mL of vaccine, while Beta and Gamma have been exposed to the virus but are not yet sick and only need 50 mL of vaccine each as a prophylactic. Suppose further that, as it happens, there is exactly 100 mL of vaccine available. This would be enough to cure either Alpha alone or both of Beta and Gamma. Assume that giving 50 mL to Alpha will produce no benefit whatsoever; it is simply not an effective dose for her. How should one distribute the vaccine in such circumstances? (Saunders, 2018, p. 68)

How, indeed, should one distribute the vaccine in such circumstances? I think that the question that Saunders poses will have different correct answers depending on who the “one” is and in what capacity and with what authority that one is distributing the vaccine. It will matter, too, what we interpret the proper purpose of the distribution of the vaccine to be. It will matter, too, to whom the vaccine is being distributed. Hence, my response to the question is not quite as straightforward as he seems to suggest.

Initially, Saunders puts forward three options. None of them, it appears to him, comfortably fits my position. He writes:

Though he does not discuss such cases, McLachlan (2012, p. 318) proposes “the random selection of names.” That is, we could put the three names into a hat and draw one out to decide who gets vaccine, thereby giving everyone a one-in-three chance of getting an effective dose. However, note that if either Beta or Gamma is drawn, there is still enough vaccine left to treat the other. Thus, it would be possible to treat both Beta and Gamma, should either one’s name be drawn, but this means that they each effectively have twice the chance of Alpha. (Saunders, 2018, p. 68)

That is his Option 1 and, according to Saunders, “if the aim is to give all citizens an equal chance of getting vaccine, in the name of impartiality, then this policy is no good” (Saunders, 2018, p. 68). This leads to Option 2, in which “any leftover vaccine should be wasted. If Beta is selected by the lottery, then Beta—and Beta alone—should be vaccinated. Though there is also enough vaccine leftover for Gamma, it would be unfair to vaccinate her too, for she was in the lottery and lost” (Saunders, 2018, p. 68).

However, although this gives each an equal chance of getting vaccine, it has the clear objection of wastefulness. This leads to Option 3. Saunders writes: “Once it is realized that Beta and Gamma can be treated together, they could be given a single lottery ticket. Thus, we can increase everyone’s chance of getting vaccinated (from 33 percent to 50 percent) and avoid waste at the same time. This proposal gives everyone the greatest equal chance of vaccination” (Saunders, 2018, p. 69). He continues:

Thus, it seems likely that he would favour Option 3 over either Option 1 (which gives some better chances than others) or Option 2 (which, by being wasteful, is worse for everyone).

If this is McLachlan’s position, this is interesting, since this has not been a popular solution to the analogous “numbers problem” posed by John Taurek (1977), in which a rescuer must choose between saving one person or two others. While consequentialists take the solution to be obvious (save the greater number), most nonconsequentialists who have considered this problem also think that numbers matter in some way.... Hence, if this is McLachlan’s position, it is not a popular one, even among nonconsequentialists. (Saunders, 2018, p. 70)

Option 1

Name drawn	Probability	Outcome
Alpha	1/3	Alpha is treated. No vaccine left over.
Beta	1/3	Beta is treated. Leftover vaccine used to treat Gamma too.
Gamma	1/3	Gamma is treated. Leftover vaccine used to treat Beta too.

Option 2

Name drawn	Probability	Outcome
Alpha	1/3	Alpha is treated. No vaccine left over.
Beta	1/3	Beta is treated. Leftover vaccine is wasted.
Gamma	1/3	Gamma is treated. Leftover vaccine is wasted.

Option 3

Name(s) drawn	Probability	Outcome
Alpha	1/2	Alpha is treated. No vaccine left over.
Beta and Gamma	1/2	Beta and Gamma are treated. No vaccine left over.

I think that my position can accommodate such objections to it as Saunders's examples here and elsewhere might seem to raise. He concludes that "the requirements of impartiality are hotly contested. Moreover, one cannot simply assume that impartial treatment will result in people getting equal chances of vaccine (as McLachlan does)... We might arrive at unequal chances, without showing any partiality for particular individuals" (Saunders, 2018, p. 71).

I certainly do not think that impartial treatment will always lead to equal chances of getting vaccine. Only under some particular assumed circumstances do I propose that, on the grounds of impartiality, citizens should be given equal chances of getting (a vaguely specified quantity of) vaccine (Wardrope, 2012; McLachlan, 2015). I agree very strongly that what counts as impartial treatment is a matter of interpretation and thereby of potential, reasonable contention. One such crucial point of contention between Saunders and me pertains to what we should suppose that the state would be responsible for distributing in an envisaged vaccine lottery. I think it would be responsible for distributing the chances of receiving an effective dose of vaccine. Saunders thinks that it would be responsible for distributing the chances of survival.

I did not make clear what I meant by an "effective dose of vaccine." I did not mean something sufficient to cure the recipient of illness if the person were ill and to prevent illness if the person were not yet ill from the infection in question. A standard dose of vaccine, rather than freedom from illness, was the prize in my supposed lottery. A scarce amount of vaccine might, in principle, be divided equally into tiny quantities and distributed equally to every citizen where that meagre dose would be totally ineffective for each recipient. By an adequate dose I meant a quantity that was likely to be effective for most, even if not all, people. I assumed that the vaccine would be distributed in standard doses and that it was not known or knowable in advance what particular amount any particular person might need in order for the vaccine to be an effective prophylactic for that particular person. For some people, the standard dose might be more than required; for some others, it might turn out to be less. Should the standard dose be such that the number of people who can receive what is likely to be an effective dose for them is maximized or should the standard dose be such that the number of people who receive a dose that is likely to be ineffective for them is minimized? An explicit or implicit political decision along such lines will, I imagine, be required. The proposed lottery gives citizens a chance of receiving such a standard dose.

Saunders is wrong to suppose that I would prefer Option 3 to either Option 2 or Option 1. He would be wrong, too, to suppose that I would prefer Option 2 to Option 1. Despite what Saunders asserts, it is not the case that, given my views, Option 1 is "no good."

How do I think one should distribute the vaccine in the circumstances that Saunders specifies? What are the salient features of the circumstances? To whom or to what does the vaccine belong? Who or what is the "one" who does the distri-

buting? In what capacity does the distributor distribute it? What is the vaccine being distributed as? For instance, is it distributed as a prophylactic, as a cure, or as that which might mitigate the symptoms of illness? In what capacity are Alpha, Beta, and Gamma potential recipients of the vaccine (McLachlan and Swales, 1999)?

Suppose that the vaccine is the property of a rich philanthropist who, as a private citizen, has financed the research, testing, and manufacture of the vaccine. Suppose that the philanthropist wants to distribute the vaccine as a gift, as an act of charity. Such a person is not under the moral constraint of a moral duty to treat the three people impartially or fairly. Whether and how the vaccine is distributed among these three people is his own business. They do not have a moral right to impartial treatment from him with regard to its distribution. Suppose that he, Beta, and Gamma are devout Christian fundamentalists. Suppose that Alpha is an atheist. He might think that it would be better if Beta and Gamma were to die of the virus rather than Alpha. He might choose to give Alpha the vaccine in order to extend her time to seek and accept salvation. There again, he might choose to refuse to give the vaccine to Alpha because he sincerely believes and has reason to think but no evidence to prove that Alpha is a child molester. This might be very unfair to Alpha. Nonetheless, the philanthropist is, in my view, morally entitled to make that choice.

If the vaccine is to be distributed by an agent of the state as the property of the state, and if Alpha, Beta, and Gamma are all citizens of that state, the distributor is under moral constraints that are dictated by that role and of which the rich philanthropist, as a private citizen, is free. For instance, discrimination on the basis of the religious beliefs of potential recipients or on the basis of the expected fate of their eternal souls would be unjust. It is not the business of an agent of the state to know or speculate about the religious beliefs of citizens or the fate of their eternal souls and far less, to discriminate between them on the basis of such speculation or knowledge. Even were such discrimination to be considered to be impartial, it would not be a difference in treatment that was grounded on a morally relevant difference, a difference that was appropriate for the agent to act upon. Nonetheless, there could be various policies and procedures that would be proper, impartial, and just. What agency or agent of the state is assumed to be in action in Saunders's envisaged situation? If the agency or agent has sufficient discretion and authority, my favoured response is to say is that, unlike Beta and Gamma, Alpha is ill and needs medical treatment. She should be admitted as a patient, if she is not one already, and treated as such. She should be given the 100 mL of vaccine that she needs. This is clearly in accordance with impartial treatment since, if either Beta or Gamma were similarly ill as she now is and if she were similarly well as they now are, either Beta or Gamma would be given the medication. A vaccine as a prophylactic is not given as a medical treatment of patients as patients but as treatment of citizens as citizens in order to try to prevent their becoming patients. As things stand in the example, Beta and Gamma are not ill. Beta and Gamma might or might not become ill, but they are not ill. Alpha is ill. She should get the vaccine as medical treatment.

However, suppose that the role of the agency or agent is specifically restricted in the circumstances to the distribution of vaccine as a prophylactic. Not all needs of potential recipients should, if known, be taken into account by the state-authorized distributor of the vaccine. They are not necessarily a relevant consideration. For instance, it might be the case that Beta needs £5000 to pay off a loan shark who has threatened to kill her if she does not provide the money by the end of the week. It might be the case the Gamma will commit suicide if his estranged lover does not return to him soon. These needs, if unmet, might render the administering of the vaccine to them useless. It does not follow that the distributor of the vaccine should take these needs into account when allocating it, nor that he should try to meet them. However, a private rich philanthropist might properly and laudably choose to do so.

Option 1 is not, as Saunders describes it, “no good.” It is a viable, impartial policy. What is distributed by the lottery is not bespoke amounts of vaccine that are sufficient to provide each winner with what he or she happens to need to survive the pandemic or to avoid dying from influenza. What is distributed is specific doses of vaccine. If 100 mL were the standard dose, to draw a name out of the hat and give the vaccine to whomever bore that name would be an appropriate impartial way of distributing the vaccine. If 50 mL were the standard dose, it would be appropriate impartial treatment to draw a name out of the hat and give a dose of vaccine to everyone except the person with that name.

There are two variants of Option 1, 1a and 1b, where, respectively, the doses are 100 mL and 50 mL.

Option 1a

Name drawn	Probability	Outcome
Alpha	1/3	Alpha is treated. No vaccine left over.
Beta	1/3	Beta is treated. No vaccine left over.
Gamma	1/3	Gamma is treated. No vaccine left over.

Option 1b

Name drawn	Probability	Outcome
Alpha	2/3	Alpha is treated. No vaccine left over.
Beta	2/3	Beta is treated. No vaccine left over.
Gamma	2/3	Gamma is treated. No vaccine left over.

Consider, firstly, Option 1a. One name is drawn from the hat and the vaccine, one hundred millilitres, are given to whomever bears that name. If Beta or Gamma receives the vaccine, he or she might still fail to survive if the debt is not paid

or the lover does not return. Secondly, consider Option 1b. One name is drawn from the hat—in this case, the name of the loser—and a dose of vaccine, 50 mL, is given to each of the other two people. In this instance, all might still perish since 50 mL is not sufficient to cure Alpha of the illness. However, in both situations, 1a and 1b, all three people have an equal chance of receiving a dose of vaccine. They all, obviously, have twice the chance of receiving a dose when there are two doses of 50 mL rather than one dose of 100 mL. But, whatever the dosage, they all have an equal change of receiving a dose. It is misleading to say that Beta and Gamma “each effectively have twice the chance of Alpha.”

It is not the business of the distributor of the vaccine to act upon the information pertaining to the personal health and medical features of the individual people concerned. Hence, it might well come about that Alpha is allocated 50 mL of the vaccine, which will not be of any use to her. Beta or Gamma might be given 50 mL more than they have any need for. So be it. Morally optimal rules and procedures can have what are—or would be, if considered in isolation—morally undesirable outcomes. It would not be the business of a member of a jury to take into consideration information not presented in the court about prior accusations made against and prior convictions of the defendant. Adherence to this good rule can have the outcome that guilty people can remain unpunished and free to reoffend.

The particular rights and duties of the role of agents of the state for the distribution of vaccine direct and constrain what counts as the proper and impartial exercise of that role. For instance, as a political rather than as a medical decision, it might be decreed that the vaccine in question should be distributed as a prophylactic in units of 50 mL only to those citizens who are not infected by the virus. How should one distribute the vaccine in such circumstances? Alpha has special needs that should be met. However, it is not the responsibility of the distributor of this vaccine to meet them or to consider them. Beta should be given one dose and Gamma the other one.

(B) Further Options

Saunders considers “Further Options” with regard to unequal needs. He asks us to

continue to assume that we have 100 mL of vaccine, which Alpha needs all of, but which is also enough to treat both Beta and Gamma. Since there is enough for Beta and Gamma, there is at least a *prima facie* case for treating this as two 50 mL doses of vaccine, though it happens that Alpha needs a double dose. It is not obvious that Alpha’s chances of getting a double dose should be equal to Beta’s chances of getting one dose. We might, instead, implement a two-stage procedure. (Saunders, 2018, p. 71)

He presents six options and indicates various variables for each of them: the chances of survival of the pandemic for each of the three people; whether or not these chances are equal; whether or not the policy is wasteful (i.e., whether all

the vaccine is used effectively); and the expected deaths. These options are as follows:

Option 4. Allocate the first dose by lottery, giving each person a $1/3$ chance of receiving it. Then allocate the second dose by lottery, between those still in need. This means, if either Beta or Gamma won the first lottery, they would no longer be in need and the second lottery would be 50/50 between the remaining two. However, if Alpha won the first lottery, she would still be in need, so the second lottery would also give each of the three a $1/3$ chance to receive the second dose....

Option 5. Allocate the first dose by lottery, giving each person a $1/3$ chance of receiving it. If Alpha wins the first dose, then allocate the second dose by lottery, giving each person a $1/3$ chance of receiving it. However, if Beta wins the first lottery, give the second dose to Gamma, since it is no use to Alpha. And, conversely, if Gamma wins the first lottery, give the second dose to Beta....

Option 6. Allocate the first dose by lottery, giving each person a $1/3$ chance of receiving it. If Beta wins the first dose, then give the second to Gamma, and vice versa. So far, this is the same as Option 5 but, if Alpha wins the first dose, then give her the second dose too, in order to avoid waste. (This is Option 1 from the previous section.)...

Option 7. Since giving the first dose to Beta will mean giving the second to Gamma, and vice versa, they can effectively “share” chances....Thus, allocate both doses together, giving a 50 percent chance to Alpha and a 50 percent chance to Beta and Gamma. (This is Option 3 from the previous section.) (Saunders, 2018, p. 71–72)

To this he adds two other options, one which maximizes the number of lives saved by dispensing with a lottery and giving the vaccine directly to Beta and Gamma and another which gives an absolute priority to equality by refraining from treating anyone. These are, respectively, **Option 8** and **Option 9**.

He outlines and compares these policies schematically as follows: (Saunders, 2018, p. 73)

Comparison of Options 4–9

Policy	A's chance	B's chance	C's chance	Inequality?	Wasteful? (Probability)	Expected deaths
Option 4	2/18	11/18	11/18	Y	Y (5/9)	1.67
Option 5	2/18	14/18	14/18	Y	Y (2/9)	1.33
Option 6	6/18	12/18	12/18	Y	N	1.33
Option 7	9/18	9/18	9/18	N	N	1.5
Option 8	0/18	18/18	18/18	Y	N	1
Option 9	0/18	0/18	0/18	N	Y (1)	3

Suppose, again, that a rich altruist acting in the capacity of a private citizen were the potential distributor of the vaccine. I think that any and all of the options would be morally permissible ones for such a person to choose. Such a person is not morally obliged to act impartially. Such a person might reasonably hold various opinions about what the morally best decision to make is. Such a person, in my view, is not morally obliged to choose to make what that person considers to be the morally very best decision is. A decision that is, morally, good enough will, morally, suffice. We are not morally obliged to act solely upon morally relevant reasons and impulses. Prejudices, subjective preferences, and nonrational and, even, irrational factors can have a morally permissible place in the decisions of private individuals.

A rich altruist might be interested to know what age Alpha, Beta, and Gamma were and whether or not they had dependent children and, if so, how many they had. A rich altruist might want to know whether or not they were single parents. Whether and, if so, how the rich altruist was influenced in making a decision about how to distribute his vaccine in the light of such information is the business of the rich altruist. The rich altruist is not answerable to us regarding how he chooses to direct his charity.

It would be morally permissible and even, I would argue, morally obligatory for a rich philanthropist to give preferential treatment to one of the potential recipients if the potential recipients happened to be the biological offspring of the philanthropist. It would not be morally permissible for the distributor of state-owned vaccine to discriminate on such a basis. Similarly, it would be morally permissible even if not morally obligatory for the rich philanthropist to choose to give vaccine to none of the three people, but to give it, instead, to people who lived in a different, poorer country. That might be the morally best outcome of the allocation and administration of the vaccine, but it is not an option that is morally permissible for agents or agencies of the state of which Alpha, Beta, and Gamma are citizens are free to adopt. They are morally obliged to give the trio and all their other citizens priority over noncitizens. Again, the morality of actions does not pertain solely to their consequences. It matters how and why particular consequences are brought about and it matters who brings them about.

Suppose that Alpha's death would be postponed by one year by the vaccine and that the lives of Beta and Gamma would each be postponed by three years were they to share it. Suppose, too, that Alpha were twenty and that Beta and Gamma were both eighty years of age. A rich philanthropist might think that it would be far fairer to give the vaccine to Alpha rather than to Beta and Gamma on the grounds that she has only lived a short life so far while the other two have lived comparatively long ones (Parfit, 1997; Hughes, 2020). The outcome that Alpha is treated might well be morally preferable to the outcome that Beta and Gamma are treated. The rich philanthropist might, not improperly, act upon that preference. However, it does not follow that it would be morally permissible for the state, its agents, and its agencies to do so.

Saunders writes:

If our *only* concern were fairness or equality, then we might choose Option 9, though it is highly suboptimal (everyone dies). Assuming that we wish to save lives, we should seek an option that saves as many lives as possible consistent with impartial treatment. McLachlan (2012, p. 318) agrees with this, though he thinks impartial treatment precludes Option 8. As we saw in the previous section, McLachlan might endorse what is here Option 7, since this gives each person the greatest chance of receiving an effective dose of vaccine compatible with everyone else receiving the same chance. (Saunders, 2018, p. 74)

I agree that, *ceteris paribus*, it is morally desirable for the state to try to save people's lives. However, it does not follow that I would agree that the government should always try to save "as many lives as possible consistent with impartial treatment" (Douglas, 2012). All of us will die sooner or later. It is more accurate to talk of the postponement of our deaths and the prevention of our deaths from particular causes at particular times than the saving of our lives. It is better that more rather than fewer deaths are postponed, *ceteris paribus*. However, we know that other things are not always equal. The deaths of some people are postponed for a long time. The deaths of some people are postponed for a very short time. *Ceteris paribus*, it is morally desirable for the state to try to save years of life. It is far from obvious that the state should try to save lives rather than years of life, *ceteris paribus*.

Suppose that a dose of vaccine given to Alpha would be expected to postpone her death by eleven years and that the same dose of vaccine divided between Beta and Gamma would be expected to postpone their deaths by five years each. What should the state prefer? Should it have a preference between the number of deaths that would be postponed and the number of years of life that would be saved by the postponement? Should the state strive to save lives—to minimize the number of deaths—rather than to maximize the number of years of life that are saved? It is not clear to me why it should or that it should.

I do not endorse Option 7, just as I do not endorse Option 3. However morally desirable the outcome of that policy might be in the particular scenario that Saunders depicts, it is not clear to me that it represents conformity to a rule that an agent or agency of the state could impartially apply in a lottery to distribute scarce vaccine in a pandemic. It is reliant upon knowledge of particular features of the potential recipients that such agents or agencies could not be expected to have, such as whether or not a particular person who had been in contact with another infected person would become ill in the absence of a dose of vaccine, whether or not that illness would be fatal, and precisely how much vaccine particular people would require to resist illness if they came into contact with the virus. Furthermore, even if they could have such information, it is not clear to me that they should act upon it. Different needs can be a relevant reason for giving different treatment, but not all different needs are a relevant reason for agents and agencies of the state to act upon. I suggest that it would be their responsibility to distribute doses of vaccine impartially and, if there were not enough vaccine in a particular time period, to distribute the chances of receiving a dose impartially.

Would I rule out Option 8—the allocation of the vaccine to Beta and Gamma directly without a lottery—on the grounds that it is not impartial? As I have previously indicated, I have a preference for the allocation of the vaccine to Alpha as curative medical treatment because she needs it rather than to any of them who are not ill as a prophylactic. Nonetheless, if a political decision were made that the scarce vaccine should be distributed solely as a prophylactic to those who were not yet infected—and, let us presume, that other treatment should be provided for those who were ill—Option 8 would seem to me to be a possibly impartial and acceptable policy.

IMPARTIALITY AND FECUNDITY

Saunders asks us to consider a particular situation where a benefit to one person can result, indirectly, in a benefit to other people (Wardrope, 2012; McLachlan, 2015). It is an assumption of the supposed scenario he depicts that the chances of survival from the pandemic are the same as the chances of receiving a dose of the distributed vaccine. He writes:

Suppose we have ten individuals, labelled Alpha through Kappa, and two doses of vaccine. With no relevant differences between individuals, it would seem reasonable to allocate the vaccine randomly, giving an equal chance (20 percent) to each individual. But what if Alpha works in the manufacture of vaccine? Let us suppose that if she receives one of the two initial doses, whether as a result of a lottery or not, then she can produce two further doses, which can be randomly allocated among the remaining individuals. Should this alter our distribution and, if so, how? (Saunders, 2018, p. 75)

To allocate the vaccines among all ten people on the basis of two random draws is Option 10. Saunders shows that if we weight the draw in favour of Alpha, increasing the probability that she will survive and produce the additional doses, we can increase everyone's chances of survival, which become equal at 25 percent. This is Option 11. However, Saunders shows that we can increase the chances of survival of everyone even more by the following policy, Option 12:

Give the first dose to Alpha, without a lottery. Allocate the other of the initial doses, and the two extra doses produced by Alpha, by lottery, giving each of the other nine individuals an equal chance. Allocating three doses between nine people means that each has a one in three (33.33 percent) chance of receiving a dose of vaccine. (Saunders, 2018, p. 77)

Saunders favours Option 12. He writes:

Alpha enjoys a larger share of the benefit than the other nine. The result is that we depart from the equality of Option 11. In this case, Alpha receives the vaccine for certain, which makes her much more likely to benefit than anyone else. The fact that Alpha's survival is instrumentally useful does not give her any greater claim to the vaccine though. Thus, while it may be rational for all involved to consent to Alpha receiving one of the initial doses, it is nonetheless unfair....It is an unfairness that we should almost certainly tolerate, since it increases everyone's chances of receiving vaccine....If we are prepared to compromise equality for the sake of efficiency by allocating the vaccine by lottery in the first place, rather than giving it to no one, then we ought also to prioritize efficiency here. (Saunders, 2018, p. 77)

Saunders suggests that his preference for Option 12 is not merely a subjective, personal one. He seems to think too that, specifically, both the distributors and potential recipients of the vaccine should prefer Option 12. He writes:

I would suggest that, if we find ourselves in these circumstances, we ought to favour Option 12. If we take the perspective of those involved, then what they care about is maximizing their own chances of survival. They ought not to care what chances others have, except insofar as those chances affect their own chances....To reduce everyone's chances for the sake of equality would be to make a fetish of equality, neglecting the fact that the only reason to care about equality is that we are distributing a good and hence that people should prefer inequality if it means more of that good for all. (Saunders, 2018, p. 78)

According to Saunders, my attitude to this policy is hard to predict. He writes:

On the one hand, McLachlan is clearly committed to the value of equality, even when it results in more deaths (McLachlan, 2015, p. 193).

On the other hand, his case for giving everyone equal chances is that the state ought to treat its citizens “the same in relevant respects unless there are relevant reasons for treating them differently” (McLachlan, 2012, p. 317, emphasis added). When he introduces priority for front-line healthcare workers, he refers to there being “at least one relevant reason for treating some citizens differently” (McLachlan, 2012, p. 318), viz. the fact that their occupation puts them in danger. This reason does not extend to those whose work involves production of vaccine who, I assume, are at no greater risk of infection than anyone else. (Saunders, 2018, p. 78)

I am not “committed to the value of equality” (McLachlan, 2000). Apart from anything else, treatment that is equal in some respects will be unequal in other respects. We cannot use the notion of equality to tell us what respects should matter. Moreover, equal treatment can produce outcomes that are unequal. In order to produce equal circumstances for particular people, we would often have to treat them unequally. I consider equality and inequality to be in themselves morally neutral terms.

Notice that I did not suggest that frontline healthcare workers should be given priority because of the danger of being infected pertaining to the nature of their work. I suggested that the state and its agencies have a particular duty of care towards those people that they employ as healthcare workers to do that which puts them in danger of being infected, a duty that it does not have towards other people—including, for instance, other frontline healthcare workers whom it does not employ and, for instance, sex workers and those who work in shops—who might be equally or even more at risk of infection (Symons, Matthews, and Tobin, 2021).

Saunders also writes:

Showing impartial (or equal) concern for everyone does not necessarily require giving them equal chances of anything. It may be that equal concern leads to unequal outcomes. We might read McLachlan’s proposal to prioritize healthcare workers as a tacit admission of this point, though it is not clear to me whether he takes this to be compatible with impartial treatment or a justifiable departure from impartiality. (Saunders, 2018, p. 74)

Prioritization of healthcare workers in the distribution of the vaccine as a prophylactic is, in my view, compatible with impartiality. The state has a duty of impartiality towards citizens as citizens among its various other duties, one of which is a duty of care towards those whom it employs to do particular jobs. Where the job of the provision of particular sorts of health care involves a particular risk of catching influenza, to give its employees preferential treatment in the allocation of scarce influenza vaccine is an appropriate way of trying to discharge that duty of care as an employer. It is compatible with fulfilling its duty to act impar-

tially towards citizens since there is a morally relevant difference between citizens as citizens and these particular providers of healthcare, some but not all of whom will also be citizens.

If Alpha works in the manufacture of vaccine and her being vaccinated would result in the availability of more vaccines than if she would not, “should this alter our distribution and, if so, how?” As usual, I would say that whether and how this should alter our distribution depends, among other possible things, on who “we” are, in what capacity “we” act, to whom the vaccine will be distributed, and for what intended purpose or purposes. If “we” are agents or an agency of the state, it might well be an appropriate and properly impartial policy to treat those who are indispensable to the production of the vaccine and who are not readily replaceable differently from other people and to give them priority in the distribution of scarce vaccine that is presently available. This is vaguely similar to the appropriate and properly impartial policies of treating patients differently in some respects from people who are merely at present potential patients and of treating university students differently from applicants and potential applicants to university. The state can treat citizens impartially as citizens, students impartially as students, and patients impartially as patients while treating students and patients differently from other citizens in some respects. Indeed, impartiality requires such different treatment in some respects. The state should treat, for instance, applicants to join the army impartially. It should treat also soldiers impartially but should not, generally, treat applicants as it treats soldiers. Similarly, the state might treat all citizens impartially as citizens and all essential workers in the production of vaccine impartially as essential workers in the production of vaccine while treating those workers differently from citizens who are not essential workers in the production of vaccine.

Hence, I have no reluctance to support the adoption of Option 12. Such support does not clash with my proposal to distribute vaccine to citizens on the basis of a blind lottery in other ones. With Saunders’s assertion that “equal consideration of everyone does not necessarily mean giving everyone equal chances” I agree completely. Only in some circumstances would an equiprobable lottery or a lottery of any sort be an appropriate way to distribute goods or services by the state, its agents, and agencies.

Saunders writes:

While...McLachlan may be right to reject a policy seeking to maximize the number of lives saved, on the grounds that this will be unfair to some, it does not follow that we should give everyone equal chances. Impartial consideration is compatible with taking efficiency into account, even if the results are contrary to strict equality. (Saunders, 2018, p. 67)

Notice that I did not mention fairness or unfairness in my discussion of this topic. I talked, rather, of justice and injustice (McLachlan, 2002; McLachlan, 2005; McLachlan, 2021; Rawls, 1985). However, I agree entirely that impartial treatment does not imply equal treatment.

SAUNDERS, IMPARTIALITY, AND RAWLS

One might, as I am, be favourably disposed towards Option 12 and its adoption by the state without accepting the Rawlsian basis that Saunders puts forward to support it (Rawls, 1999). Apart from anything else, there is a difference between thinking that, morally, the world would be a better place if Option 12 were to come about and thinking that any particular agent or agency is morally permitted, far less morally obliged, to try to promulgate or implement any policy that would bring that outcome about. Our moral duties might even lead us in a contrary direction.

Saunders writes:

This situation is similar to Rawls's famous original position....Here, Rawls assumes that parties are concerned only with their absolute position and not how they stand relative to others. Thus, he proposes that they would reject strict egalitarianism in favour of the difference principle, which permits inequalities that benefit all....My proposal is an application of similar reasoning to the problem at hand. That is, I suggest that vaccine ought to be distributed equally unless an unequal distribution benefits everyone. This policy satisfies the requirement to treat everyone impartially. It explains why we should prefer allocating effective doses of vaccine by lottery to giving everyone an equal but ineffective share of the vaccine. And it also tells us that, if prioritizing Alpha improves everyone's chances of survival, this is what we should do. (Saunders, 2018, p. 78)

It is not clear how, in practice, Saunders's principle "that vaccine ought to be distributed equally unless an unequal distribution benefits everyone" could apply or is meant to apply. We can, as an exercise of philosophy, suppose that Alpha and other hypothetical people will need such-and-such medication to survive and will react in particular specified ways to particular doses of vaccine when we formulate our abstract examples. However, actual distributors of vaccine can never have such knowledge. For instance, in practice, no matter how thoroughly a vaccine is tested before it is used, some people might have unfortunate, even fatal, reactions to it. It can never be known for certain that this will not happen. It is reasonable to suspect that with any vaccine against, for instance, COVID-19, if we give enough people a dose, at least someone or other will not only significantly fail to benefit from it but will, to some or other extent, suffer.³ How, if at all, should this consideration influence our distribution policy? Should we always distribute such vaccine equally? If the suggested justification of policies regarding such things as the distribution of vaccine is a sort of consequence that

could not be known prior to the implementation of the policies, it is not clear that the suggested justification is of practical usefulness to agents and agencies who have the responsibility of distributing vaccine.

Notice, too, the complication of a time gap in the implementation of Option 12. If we give a dose of vaccine to Alpha and allocate the remaining one by lottery rather than allocating two by lottery initially, there is a potential loser—someone who might have won the dose that was given to Alpha might catch the disease and die before Alpha gets round to producing the replacement and the additional dose. Furthermore, there is a risk that she will not actually produce the two doses. She might die in a road accident. She might resign from her job as, for instance, a consequence of the unexpected inheritance of a fortune. It cannot be known in advance that Saunders's principle will be satisfied by any particular application of Option 12.

Furthermore, when distributions are just and unequal, their justice is not dependent on whether or not they conform to Saunders's suggested principle. For instance, it is, I suggest, just that motor-car drivers are not all treated the same with regard to the possibility of parking without legal penalty in public places and that special treatment is given to those who have particular disabilities and impairments. They have a particular need which other drivers lack for available and convenient parking spots. This is a morally relevant difference. It is proportionate to the unequal treatment. Whether or not all people or all car drivers benefit from this arrangement as compared to one of strict equality is irrelevant to the justice or injustice of this particular arrangement.

CONCLUSION

Saunders writes: "My position is that the state owes citizens something like equal consideration or treatment as equals, but this need not require giving them equal amounts of anything, be it vaccine, chances of vaccine, etc." (Saunders, 2018, p. 81, note 5). He concludes his paper by, similarly, saying: "Thus, while I believe that policymakers should show equal concern for all citizens, I suggest that this may permit, or even require, departures from equality in the distribution of (chances of getting) vaccine" (Saunders, 2018, p. 80). With all this I totally agree. My proposed nonconsequentialist policy for the ethical distribution of scarce vaccine in the face of an influenza pandemic reflects this agreement and withstands Saunders's criticism of it.

I do not regard lotteries as such as moral ends in themselves but, in some circumstances, for some ends, as appropriate means. It is not my view, that, in the general run of things, medical treatment should be distributed by the NHS on the basis of a lottery system. For instance, I do not suggest that at the reception of accident and emergency departments, prospective patients should be given what functions as a lottery ticket on arrival. The combination of the rule of first come, first served and a triage system that takes account of the extent and

urgency of the medical needs of the relevant particular individuals serve to give, in the circumstances, a practical expression of the principle of impartiality.

At the start of this paper was highlighted Saunders's claim that "the proposal to give 'equal chances' to everyone is not as simple or as appealing as it first appears. Any policy will have to be sensitive to differences, in need and productive capacity, between people" (Saunders, 2018, p. 80). As I have shown, we cannot say what people must be sensitive to or what sort of principles must guide their distribution of scarce goods or services without considering the nature and context of the particular roles in which these particular people act and the nature and function of that which they distribute. When people act not as private individual citizens but, rather, in the role of officials or agents of the state, they act within particular spheres of authority and acquire particular moral and legal rights and duties along with their occupancy of such roles. In order to fulfil their moral duty to treat citizens impartially, agents and agencies of the state are sometimes morally required to treat particular citizens differently from others. This can be—but will not necessarily be—on the grounds of differences pertaining to their particular needs or abilities. This is not the same as treating some citizens differently from others on the grounds of differences in the needs or abilities, on average, of the members of the particular categories they belong to. Additionally, the notion of "need" does not fit comfortably with the distribution of vaccine to healthy people as a prophylactic.

NOTES

- ¹ See UK Department of Health, *Influenza Pandemic Contingency Plan* (Department of Health Publications, 2005), p. 13. See also *Responding to Pandemic Influenza: The Ethical Framework for Policy and Planning* (Department of Health Publications, 2007) and UK Department of Health, *UK Influenza Pandemic Preparedness Strategy* (London: Department of Health, 2011).
- ² As I have written elsewhere: “there is more to politics than ethics and more to ethics than justice. Justice is a virtue but it is not the only one and one that can conflict with other ones. A public library system, for instance, which was completely just but where justice was its only virtue would be a poor one and one not always to be preferred to other systems which had more and better books, less justly – perhaps even unjustly – distributed.” (McLachlan, 2005, p. 13)
- ³ In *The Herald*, June 17, 2021, in an article entitled “Covid Vaccine: Four Dead Due to Adverse Effects in Scotland,” Sara Paciaroni writes: “A report from National Records of Scotland (NRS) said that by the end of May there had been four deaths where the ‘underlying cause of death was adverse effects of Covid-19 vaccines.’ ”

REFERENCES

Colgrove, James, “Immunization and Ethics: Beneficence, Coercion, Public Health, and the State”, in Anna C. Mastroianni, Jeffrey P. Kahn and Nancy E. Kass (eds.), *The Oxford Handbook of Public Health Ethics*, Oxford, Oxford University Press, 2019, p. 435-447.

Douglas, Thomas, “The Concise Argument”, *Journal of Medical Ethics*, vol. 38, no. 5, 2012, p. 257.

Hughes, Jonathan A., “Lockdown and Levelling Down: Why Savulescu and Cameron Are Mistaken about Selective Isolation of the Elderly”, *Journal of Medical Ethics*, vol. 46, no. 11, 2020, p. 722-723.

Jansen, Lynn A. and Steven Wall, “Weighted Lotteries and the Allocation of Scarce Medications for Covid-19”, *The Hastings Center Report*, vol. 51, no. 1, 2021, p. 39-46.

Emanuel, Ezekiel J. and Alan Wertheimer, “Who Should Get Influenza Vaccine When Not All Can?”, *Science*, vol. 312, no. 5775, 2006, p. 854-855.

McLachlan, Hugh V., “Scotland, Social Justice, Health and Inequality”, in Christopher Nottingham (ed.), *The N.H.S. in Scotland: The Legacy of the Past and the Prospect of the Future*, Farnham, Ashgate Press, 2000, p. 106-123.

———, “Tobacco, Taxation and Fairness”, *Journal of Medical Ethics*, vol. 28 no. 6, 2002, p. 381-383.

———, *Social Justice, Human Rights and Public Policy*, Glasgow, Humming Earth, 2005.

———, “Moral Rights to Life, both Natural and Non-natural: Reflections on James Griffin’s Account of Human Rights”, *Diametros*, vol. 26, 2010, p. 58-76.

———, “A Proposed Non-consequentialist Policy for the Ethical Distribution of Scarce Vaccination in the Face of an Influenza Pandemic”, *Journal of Medical Ethics*, vol. 38, no. 5, 2012, p. 317-318.

———, “On the Random Distribution of Scarce Doses of Vaccine in Response to the Threat of an Influenza Pandemic: A Response to Wardrope”, *Journal of Medical Ethics*, vol. 41, no. 2, 2015, p. 191-194.

———, “Exploitation, Criminalization, and Pecuniary Trade in the Organs of Living People”, *Bioethical Inquiry*, vol. 18, no. 2, 2021, p. 229-241.

McLachlan, Hugh V. and John K. Swales, “A Drunk Driver, A Sober Pedestrian and the Allocation of Tragically Scarce and Indivisible Emergency Hospital Treatment”, *Health Care Analysis*, vol. 7, no. 1, 1999, p. 5-21.

Parfit, Derek, “Equality and Priority”, *Ratio*, vol. 10, no. 3, 1997, p. 202-221.

Peterson, Martin, “The Moral Importance of Selecting People Randomly”, *Bioethics*, vol. 22, no. 6, 2008, p. 321-327.

Rawls, John, “Justice as Fairness: Political not Metaphysical”, *Philosophy and Public Affairs*, vol. 14, no. 3, 1985, p. 223-251.

Rawls, John, *A Theory of Justice*, revised edition, Oxford, Oxford University Press, 1999 [1971].

Saunders, Ben, “Equality in the Allocation of Scarce Vaccines”, *Les Ateliers de l'éthique/The Ethics Forum*, vol. 13, no. 3, 2018, p. 65-84.

Savulescu, Julian and James Cameron, “Why Lockdown of the Elderly is Not Ageist and Why Levelling Down Equality Is Wrong”, *Journal of Medical Ethics*, vol. 46, no. 11, 2020, p. 717-721.

Symons, Xavier, Steve Matthews and Bernadette Tobin, “Why Should HCWs Receive Priority Access to Vaccines in a Pandemic?”, *BMC Med Ethics*, vol. 22, no. 79, 2021, <https://doi.org/10.1186/s12910-021-00650-2>

Taurek, John M., “Should the Numbers Count?”, *Philosophy and Public Affairs*, vol. 6, no. 4, 1977, p. 293-316.

Wardrope, Alistair, “Scarce Vaccine Supplies in an Influenza Pandemic Should Not Be Distributed Randomly: A Reply to McLachlan”, *Journal of Medical Ethics*, vol. 38, no. 12, 2012, p. 765-767.

Williams, Jane H. and Angus Dawson, “Prioritising Access to Pandemic Influenza Vaccine: A Review of the Ethics Literature”, *BMC Med Ethics*, vol. 21, no. 40, 2020, <https://doi.org/10.1186/s12910-020-00477-3>

DOSSIER :

AFFECTIVE STATES, HAPPINESS, AND WELL-BEING:

INTRODUCTION

SAMUEL LEPINE

UNIVERSITÉ CLERMONT AUVERGNE, LABORATOIRE PHIER

Affective states and happiness have been the subject of much philosophical discussion over the last decades. The roles that desire and pleasure are likely to play in our being happy have received considerable attention. Intuitively, one might think that it is indeed enough to experience many pleasures or to satisfy most of our desires to be happy, but things are obviously far from being that simple. Many pleasures seem to compromise our happiness in the more-or-less long term, while the satisfaction of many desires may leave us indifferent or give way to bitter disappointments. Moreover, understanding happiness only in terms of pleasures or satisfied desires may oversimplify the picture. Our different positive emotions also probably play an important role in happiness. Whereas pleasure is often characterized as a felt sensation with a positive valence, but without intentionality, it is generally agreed that emotions have a richer evaluative aspect, notably because they have intentionality: to feel pride about an achievement, for instance, is a state that is directed at this achievement, which consists in evaluating it as a success, and that is at the same time endowed with a positive valence. From this point of view, emotions seem to play an important role in the appraisal of what is important for our happiness. More generally, the idea that people can be happy simply because they feel a lot of pleasure or because their desires are satisfied, and this independently of their overall emotional life, is implausible. The emotions that people feel, in fact, also say something about their relation to the world and about the way events affect them, and they are in this sense richer than the simple positive feeling of a pleasure or the satisfaction of a desire.

The picture becomes even more complex because philosophers have also distinguished between happiness and well-being. A typical way to understand this distinction is to say that happiness is a descriptive concept that refers to the psychological state of a person, whereas well-being is a normative concept, used to determine whether a person's life is going well or is flourishing, according to a certain standard. Haybron (2008) offers the following example, inspired by Kagan's "deceived businessman" objection (Kagan, 1994): let's imagine that George is completely satisfied with his life, that he feels fulfilled, while those close to him mock him without his knowledge. They all make him believe that they love him because he is rich, whereas they cannot stand him. Intuitively, it

seems that George is happy, but that his life does not conform to some of our expectations about what a flourishing life is supposed to be. It seems indeed conceivable that people can feel happy while their levels of well-being remain low. Theories of well-being, from this point of view, usually try to specify which things are likely to be good for an individual in themselves, in a non-instrumental way (Fletcher, 2016). Thus, things that are likely to increase an individual's well-being are said to have prudential value. While happiness thus designates the psychological state of a person, well-being will rather refer to a normative evaluation of that person's life. A crucial question is then to determine the weight to be given to happiness in the well-being of an individual. To assert that happiness is an essential component of an individual's well-being is to assert that happiness also has a prudential value.

The aim of this issue is to explore the different links between affective states, happiness, and well-being, and its focus will be on the psychological and metaethical issues that these questions raise. While some contributions focus on elucidating the nature of the affective states that contribute to happiness, others aim at a more general characterization of happiness as an emotional condition and of its relation to well-being. Some contributions also seek to define the conditions that an affective theory of well-being should satisfy to be convincing. Before going into the details of these contributions, however, I would like to map the major theories of happiness today, as well as the links that can be forged between affective states, happiness, and well-being.

The most well-known theory about happiness is probably hedonism, which consists in claiming that happiness is a positive hedonic balance of pleasures over pains. Thus, a happy individual would simply be one who, on average, experiences more pleasures than pains. On this approach, it is natural to think that pleasure also has a prudential value, and that our well-being increases because our pleasures increase on average. This theory has the advantage of explaining in a relatively simple and intuitive way why things contribute to both our happiness and our well-being: it is simply because their experience is pleasurable. This approach has many other advantages, but it has also met with a certain number of recurring objections, the main one being that pleasure does not always seem to contribute in a decisive way to our well-being. This is because, on the one hand, certain pleasures seem superficial (such as the pleasure of scratching oneself or drinking a cheap beer) and, on the other hand, because we can easily conceive of individuals who would experience many pleasures through a machine giving them the illusion of living all sorts of events, but who would not appear to us to be happy or fulfilled, notably because we would probably not want to exchange our life for theirs (Nozick, 1974).¹

The other major theory of happiness focuses on how satisfied we are about our lives. According to this approach, a person's happiness varies according to that person's level of satisfaction with his or her life as a whole. In other words, happiness depends on the judgment we are likely to make about our life, considering whether it is in line with our goals and our expectations or whether it has

a meaning for us (the priority to be given to these different ingredients is of course subject to discussion). Let us note in favour of this theory that many empirical studies consist in mobilizing this kind of judgment of satisfaction: research in psychology on “subjective well-being” (which is in fact equivalent to happiness as we have defined it), for example, often consists in asking individuals how satisfied they are with their lives on a scale of 0 to 10. This theory thus puts more emphasis on the cognitive dimension of happiness than on its felt dimension, since happiness depends here above all on a judgment that we make about our lives. Based on this kind of approach, Sumner (1996) proposed that we conceive of well-being as a form of authentic happiness. According to him, well-being would thus refer to a happiness that is both informed and autonomous. In order for my judgment of satisfaction to be reliable, indeed, it seems necessary that it take into account as much information as possible about my own life, but this judgment should also be formulated without my being under the influence of a form of conditioning or habituation to disastrous living conditions for which I will have eventually developed an adaptive preference.

Still, it is tempting to criticize this theory for placing too much emphasis on the reflective dimension of happiness (Haybron, 2008). At the very least, it seems conceivable that people could be satisfied with their lives, while having disastrous emotional experiences. Conversely, one can imagine people who would be dissatisfied with their lives without being unhappy for all that: one can regret certain career choices or the obligation to live with a smartphone, while otherwise enjoying a certain level of happiness. In order to take into account this affective dimension of happiness, Haybron puts forward the emotional-state theory, according to which happiness consists in the overall positive emotional condition of a person. This condition is characterized by various properties and, in particular, by the fact that a happy person is supposed to manifest a “propensity” to experience positive emotions and moods, which enable that person to respond favourably to the various circumstances that affect his or her life. A happy person, in this sense, is one who is not easily disturbed by small everyday troubles and who generally displays a kind of welcoming emotional attitude. Happiness, according to Haybron, would then have a prudential value insofar as it constitutes an aspect of personal fulfillment that is at least a part of well-being. This approach thus suggests an alternative to both hedonistic and life-satisfaction theories, while retaining their respective strengths. As with hedonism, the emotional-state theory bases itself on affective states, but it does so by appealing to the vast palette of our emotions and moods, not only to our pleasures. As with the life-satisfaction theory, it also emphasizes a form of satisfaction, but one which would be understood as a felt affective state rather than a judgment.

The emotional-state theory has been the subject of much debate for more than a decade, and many of the contributions in this issue are devoted to discussing some of its aspects. Thus, Haybron himself opens this issue by first reminding us of some of the central aspects of this theory, before returning to the difficulties posed by his description of happiness in dispositional terms. To be happy,

according to Haybron, implies a form of emotional resilience that can be understood as a disposition to experience positive emotions and moods over the long term, despite the occasional episodes of sadness that we may experience. In this article, Haybron attempts to clarify the metaphysical characterization of happiness as a dispositional state. He understands happiness as a concept that allows us to evaluate the functional condition of an individual: it is an indicator of that individual's level of well-being and ability to respond to the events he or she encounters. Happiness, in this sense, would not be a passing state, but rather a general state, as is the state of our health or the state of an engine. Far from being mysterious, happiness would thus consist in a certain operational state that designates the "preparatory condition" of an individual who is disposed to respond favourably to events, just as an engine can be in a condition that allows it to function well. Conversely, depression would also be a dispositional state insofar as it consists in being disposed not to go well and not to be able to respond favourably to the events we encounter. Haybron also points out that the dispositional character of happiness is presupposed by empirical research on subjective well-being, since it is precisely this background state that we refer to in order to assess our level of satisfaction with life.

Many philosophers have followed the path opened by Haybron with the emotional-state theory and developed different variants of this view. In the second article of this issue, Rossi and Tappolet propose, for example, that we conceive of happiness as the positive balance that results from our different affective states, such as our pleasures, our emotions, and our moods. What is decisive for an affective state to contribute to our happiness is its "centrality," a concept that was already crucial in Haybron's theory. According to Haybron, a state is central if it disposes us to experience other affective states. An enduring positive mood, for instance, is central if it also disposes us to experience other positive emotions. Rossi and Tappolet understand "centrality" somewhat differently: a central state, according to them, refers to what we value, where valuations should be understood as sentiments like love or hate. The different affective states that contribute to our happiness can all be understood, according to Rossi and Tappolet, as perceptive states, whose function would be to track certain evaluative properties (an emotion of fear can track the dangers that threaten what we care about, a sentiment of love can track loveable persons, etc.). If these states are correct (i.e., if we fear something that is dangerous, or love someone who is loveable), then the happiness resulting from these states would itself be correct. And well-being itself, according to Rossi and Tappolet, would then be a state of appropriate or correct happiness.

In the third contribution of this issue, Antti Kauppinen turns to the relationship between virtue and happiness, understood once again as an emotional condition. He particularly challenges the idea that the possession of virtues is necessary to be happy. Basing himself on a careful examination of neo-Aristotelian positions, Kauppinen argues that they generally defend a conception of happiness that is not credible, insofar as virtues are likely to require us to put ourselves in danger or to renounce something we desire, so they would be likely to generate a certain

number of unpleasant affective states. However, Kauppinen readily admits that virtues can be conducive to happiness, but he argues that there is no necessary causal connection between possessing a virtue and being happy. To the extent that virtue does indeed involve a concern for things that are valuable, then the virtuous agent will be likely to have positively valenced experiences as a result of this concern (the virtuous agent will seek out sincere and solid friendships, will have a secure sense of self-worth, will be less likely to suffer from anxiety or depression, etc.), and there is a strong possibility that the virtuous agent will thus be able to settle into a virtuous circle where the positive emotions and moods he or she experiences will themselves make it more likely that other pleasant emotional states will come about. Finally, Kauppinen discusses some of the implications of this approach from the point of view of well-being, contrasting it with the view defended by Rossi and Tappolet.

If the emotional-state theory proposes that we conceive of happiness as a disposition to experience positive moods and emotions, other sentimentalist approaches have characterized happiness, in a more minimalist way, as a *sui generis* emotion. Lepine first examines and discusses these approaches, before defending a characterization of happiness in terms of mood. According to him, happiness is neither an emotion nor a disposition to experience certain emotions and moods, but rather a positive mood that is an occurrent and discrete state. He thus argues that happiness is nothing other than a positive mood. To this end, he characterizes moods as prospective attitudes, whose purpose would be to appraise the more-or-less propitious character of situations with respect to the things we value. He then argues that moods have a prudential value when they are correct, and that they can contribute to our well-being when this condition is met.

Deonna and Teroni, for their part, ask not whether happiness can be understood as an emotion, but rather to what extent emotions are likely to contribute to the happiness of individuals. Since emotions all have a hedonic component that can be positive or negative, so that they are more or less pleasant or unpleasant to feel, it is indeed natural to think that they must contribute in some way to our happiness. Deonna and Teroni's claim is that emotions contribute to our happiness insofar as they constitute specific forms of engagement with the world: they are evaluative attitudes that consist mainly of feelings of preparation for action. As such, they are ways of taking pleasure or displeasure that are phenomenally translated into "types of engagement" with the objects of our emotions. But this approach raises other issues, which Deonna and Teroni go on to clarify—notably, the question of what emotions, as states of pleasure or displeasure, have in common, and the question of whether the tendencies to act that are specific to emotions are sufficient to contribute hedonically to our happiness.

In a more general way, Broi is interested in ethical hedonism, where the only thing that is finally good is pleasure. He offers an analysis of the constraints bearing on this theory based on recent research on pleasure and argues that only one of the current theories of pleasure is compatible with ethical hedonism. In

particular, he distinguishes between internalist theories (which equate pleasure with a quality of our experience) and externalist theories (which equate pleasure with an attitude that we adopt towards a type of experience). According to internalist theories, pleasure would have a final value because of its own phenomenal character, whereas externalist theories, conversely, would have it that pleasure derives its final value from the attitudes we have towards an experience (for instance because we desire it for its own sake). This second approach amounts to saying that pleasure alone would have no final value, which is clearly problematic for hedonism. After reviewing the reasons why internalist theories have generally been set aside, Broi presents some hybrid hedonist theories that emphasize both pleasure and our desire for pleasure: the final value of pleasure, according to these approaches, would be constituted by both the pleasure and our attitude toward it. In a similar vein, representationalist theories argue that pleasure derives its value from the fact that we represent it as good. In both cases, pleasure would nevertheless derive its final value from something other than itself—namely, an attitude of the subject. According to Broi, ethical hedonism is then in a bad position with respect to current theories of pleasure.

In his contribution, Stéphane Lemaire asks what kind of pro-attitude a subjectivist theory of well-being should defend. According to subjectivist approaches, indeed, well-being depends on certain subjective properties such as our desires or our preferences. But this raises the question of which attitude in particular a subjectivist theory should focus on. In order to answer this question, Lemaire reviews the different desiderata that a subjectivist theory should satisfy. These desiderata are that of extensional adequacy (the pro-attitude we consider relevant to well-being should not prevent us from recognizing that a thing or a situation could contribute to a person's well-being or ill-being), that of invariability (well-being should always refer to the same property when considering human beings, be they children or adults), and that of existence (the pro-attitude we refer to should actually exist and not be merely postulated). Lemaire then examines different kinds of theories that make desire the relevant pro-attitude, showing that they are all exposed to considerable difficulties from the point of view of the first two desiderata. He then presents an alternative hypothesis according to which the relevant pro-attitude would be a *sui generis* attitude of valuation which, he sets out to show, satisfies precisely all the desiderata weighing on subjectivist theories of well-being.

Finally, Julien Claparède-Petitpierre's contribution takes a more historical interest in the way in which well-being has been studied in the social sciences and asks how the question of well-being has been structured with regard to the analysis of social normativity. In particular, he identifies two analyses of this normativity, which he considers to be complementary. According to the first, this normativity would be "moral," and well-being would depend on the relationship between an individual's character and the moral requirements with which that individual is confronted. According to the second, this normativity would be "structural," and well-being would depend on the affective relations that the individual maintains with an environment that he or she perceives as more-or-less familiar.

The eight contributions gathered in this issue thus approach the concepts of happiness and well-being from their affective constituents. While the first four papers revolve around the concept of happiness in terms of emotional condition, the following four discuss more specifically the way in which emotions, pleasures, or valuations are likely to contribute to happiness and well-being. Undoubtedly, the reader will find many echoes from one text to another, and thus we hope that this issue will contribute to enriching the discussion of the relationships among affective states, happiness, and well-being.²

NOTES

- ¹ However, see De Brigard (2010), who suggests that our responses to Nozick's thought experiment are primarily influenced by the status quo bias.
- ² I would like to thank Julien Deonna and Stéphane Lemaire for their precious help and remarks on this introduction. I also warmly thank *The Ethics Forum* for editing this issue and in particular Jean-Philippe Royer, Guillaume Soucy, and Éliot Litalien. Finally, I would also like to thank Magalie Schor, with whom I had started to prepare this issue, but who unfortunately could not continue this project.

BIBLIOGRAPHY

De Brigard, Felipe, “If You Like It, Does It Matter If It’s Real?”, *Philosophical Psychology*, vol. 23, no. 1, 2010, p. 43–57.

Fletcher, Guy, *The Philosophy of Well-Being: An Introduction*, London, Routledge, 2016.

Haybron, Dan, *The Pursuit of Unhappiness*, Oxford, Clarendon Press, 2008.

Kagan, Shelly, “Me and My Life”, *Proceedings of the Aristotelian Society*, vol. 94, 1994, p. 309–24.

Nozick, Robert, *Anarchy, State, and Utopia*, Oxford, Basil Blackwell, 1974.

Sumner, Wayne, *Welfare, Happiness, and Ethics*, Oxford, Clarendon Press, 1996.

HAPPINESS AND THE METAPHYSICS OF AFFECT

DANIEL M. HAYBRON

PHILOSOPHY, SAINT LOUIS UNIVERSITY

ABSTRACT:

This paper introduces a category of functional conditions to address certain difficulties that have arisen in philosophical work on the nature of happiness. In earlier work, I defended an emotional state theory of happiness on which being happy consists substantially in dispositional states, such as one's propensity for a relaxed or cheerful mood. Hedonistic accounts of happiness, which reduce it to experiences of pleasure, were rejected partly on the grounds that they appear to commit a category mistake. However, the nature of this category mistake remained unclear, and the claimed dispositionality of happiness has likewise been challenged even by commentators otherwise sympathetic with an emotional state theory. Here I address these worries by sharpening the metaphysical underpinnings of the emotional state view as I have articulated it. Understanding happiness in terms of an individual's functional condition resolves these puzzles in a way that helps to explain the distinctive significance of happiness.

RÉSUMÉ :

Cet article présente une catégorie de conditions fonctionnelles pour aborder certaines difficultés qui ont surgi dans le travail philosophique sur la nature du bonheur. Dans des travaux antérieurs, j'ai défendu une théorie du bonheur comme état émotionnel selon laquelle être heureux consiste essentiellement en des états dispositionnels, tels que la propension à une humeur détendue ou ravie. Les conceptions hédonistes du bonheur, qui le réduisent à des expériences de plaisir, ont été rejetées en partie parce qu'elles semblent commettre une erreur de catégorie. Cependant, la nature de cette erreur de catégorie est restée incertaine, et la dispositionnalité revendiquée du bonheur a également été contestée même par des commentateurs par ailleurs sympathiques à une théorie de l'état émotionnel. Ici, je réponds à ces inquiétudes en précisant les fondements métaphysiques de la conception de l'état émotionnel telle que je l'ai articulée. Comprendre le bonheur en termes de condition fonctionnelle d'un individu résout ces énigmes d'une manière qui aide à expliquer la signification distinctive du bonheur.

1. INTRODUCTION¹

Here's a common enough thought about happiness: to be happy is just to feel happy. In such claims it is usually clear that "happy" is being used as a purely psychological term and not as an evaluative expression for well-being, or what Aristotle called "*eudaimonia*." No value judgment seems to be at issue here, so it can be perfectly intelligible to say things like "happy people are pathetic fools." We're just describing someone's state of mind. And yet, even in what's called the long-term psychological sense of "happiness," few philosophers conflate *being* happy with the acute emotion of *feeling* happy. At least, it is rare to find a philosophical theory of happiness, in any important sense, as amounting to nothing more than a particular emotion or mood. Rather, philosophers have focused on something to do with a person's state of mind over a more or less extended period of time. If you are happy "these days," that must be a matter of rather a lot of feelings, if it's about feelings at all. And it probably includes feelings other than the smiley-face one. So perhaps happiness is better framed in terms of *feeling good*, where this might include all manner of positive affect besides feeling happy, like a pleasant sense of tranquility. Indeed, this is something of a commonplace.

More broadly, it is entirely standard to think of happiness as a wholly *occurrent* state of mind: to be happy is a matter of what actually transpires in one's head and not, for example, how one is *disposed* to feel. Scholars sometimes frame this latter idea in terms of "dispositional happiness": a trait, such as being a happy person.² But you could possess "trait happiness" without being happy in the present; though blessed with an upbeat and laid-back nature, maybe you're having a bad week.

These are natural suppositions, but I think they are mistaken. Dispositionalism is a crucial aspect of happiness, understood not as a trait but as the relatively lasting psychological state that animates most everyday concern with happiness, including the desire to be happier that fuels a multibillion-dollar self-help industry. I have argued at length for this idea in earlier work defending an "emotional state" theory of happiness, and the broad approach has gained some uptake in the philosophical literature.³ But the basic schema of emotional state theory admits of various readings, including forms that take happiness to be entirely occurrent, or at least to be dispositional to a far lesser extent than I have claimed. For instance, perhaps happiness is purely a matter of occurrent emotions and moods, including feeling happy, serene, and energetic, and, on the negative side, feeling sad, anxious, lethargic, stressed, and so forth. (Note: I often use "happiness," like "health," as a generic term for a domain of concern that includes both positive and negative states. Accordingly, many of my examples will involve unhappiness or "unhappiness-constituting" states like anxiety.)

While I have suggested that even these states have dispositional elements, not all commentators have agreed, and there has been considerable resistance to my suggestion that some happiness-constituting states are not occurrent at all, such

as a heightened propensity for anxiety or for feeling happy.⁴ And, not surprisingly, emotional state theories have been rejected *tout court* by hedonists, who identify happiness with wholly occurrent states of pleasure, versus displeasure.⁵ So the form of emotional state theory of happiness I've defended faces skepticism from at least two directions owing to its focus on dispositionality.

This focus can indeed seem odd if the venerable doctrine of hedonism, on which happiness just *is* pleasure, serves as one's reference point in thinking about happiness. There are better ones: for starters, the popular English locution, "happy and healthy." (I do not know how prevalent equivalents are in other languages, though "*heureux et en bonne santé*" has some purchase in contemporary French.) Whatever the expression refers to, people tend to think it quite important, so that they'll often say things like "I just want my kids to be happy and healthy." This frequently reduces to "I just want my kids to be happy." If happiness alone is that important, then happiness and health must be really important. It is doubtful that in making such assertions many of us mean to be committing ourselves to the idea that these things literally exhaust the constituents of a good life. It's just that the other bits, like not being a serial killer, don't require mentioning. Moreover, the fact that "and healthy" seems not to be redundant suggests that whatever happiness is thought to be, it isn't necessarily the same thing as well-being, even if "happiness" can take on such a meaning in other contexts. What is being wished for, beyond health, is very likely something to do with mental states: happiness in the long-term psychological sense.

I think the parallel with health reveals a crucial feature of the way in which happiness is ordinarily conceived: namely, as something belonging to the category of what I'll call *functional conditions*, or "conditions" for short. Just as ascribing health to someone is saying something about that person's condition, so, I will argue, is ascribing happiness to someone. If this is right, then the dispositionality I have attributed to happiness should not be puzzling after all. For dispositionality is an essential feature of conditions; to be healthy, for instance, is at least partly to be disposed to function well, not to drop dead of a stroke, and so on. Likewise, to be happy is at least partly to be disposed to respond favourably to things in one's life, and this licenses various predictions and explanations—for example, that one will likely handle annoyances with relative aplomb. Put another way, the concept of happiness, like the concept of health, is what we might call a *condition assessment concept*. (One might alternatively refer to it as a "functional assessment" concept, a framing I owe to John Doris.)

In what follows I will, first, rehearse the essentials of the theory of happiness in question and the puzzle it raises: why think happiness is dispositional and is not merely an occurrent state? To explain this, we shall need to introduce a distinction in the metaphysics of states—namely, the idea of a condition, which I am not sure has been elaborated previously, at least in the sense intended here. To motivate the distinction and illustrate its importance in the present context, we imagine an artificial system, a security robot that monitors and responds to threats in the environment. To characterize this system in relation to its purpose,

we need concepts having to do with various aspects of its state, including its condition—for example, whether it is on high alert.

The upshot is that the emotional state theory of happiness, dispositions and all, is well motivated, capturing an important and natural aspect of a person's psychological state: that person's emotional condition. As with the security system, my emotional condition amounts to a summary evaluation of my circumstances relative to the goals that structure my makeup: are things going well for me? This underwrites the use of happiness as a convenient, nonevaluative proxy for well-being—a relatively simple way to tell how someone is doing.

The paper concludes with some reflections about the import of these points for the empirical study of personality and well-being, both of which have deficiencies owing to a failure to distinguish functional conditions from mere events in the emotional realm. A person's emotional life is not just a series of transient responses to stimuli, as is commonly supposed; like life satisfaction attitudes, it concerns the broad character of one's circumstances, and may indeed offer a better gauge of those than a person's explicit judgments. Moreover, a state's being dispositional does not suffice to make it a trait, as is widely supposed in social and personality psychology, with the result that personality measures are sometimes confounded with happiness measures. And "happiness" measures typically fail to attempt to measure key aspects of happiness, though to some extent this may be an inherent limitation of standard self-report methods.

At the core of this paper are several distinctions in the metaphysics of states and events that, to the best of my knowledge, are novel.⁶ Presumably the arguments ramify for other questions in metaphysics, but the relevant literature is vast and intricate, with considerable dispute about even basic questions. As my aim is chiefly to clarify some issues in the psychology of affect, I will try to minimize entanglement in those debates. Where that can't be avoided, it may be possible to translate my claims to accommodate different accounts of events, states, and so forth, so long as the basic relationships I've posited among different affective kinds are preserved. I regard the framework sketched below as a first approximation, not a finished product—but hopefully good enough to sustain my claims about the nature and significance of happiness.

2. BACKGROUND: THE EMOTIONAL STATE THEORY OF HAPPINESS

The theory in question, the emotional state theory of happiness, was introduced as an alternative to the then-dominant views of happiness in the long-term psychological sense, hedonistic and life satisfaction accounts. The former identifies happiness with a person's balance of pleasant versus unpleasant experiences, while the latter identifies it with a judgment-like attitude of being satisfied with one's life as a whole, or with one's life as it is these days.⁷

In its basic or default form, the emotional state account reduces happiness to a person's emotions and moods, taken together. The basic idea is that happiness

is roughly the opposite of anxiety and depression, these being prominent forms of unhappiness. The view differs from hedonism, first, in that emotions and moods are rich, deep affective states with various aspects—some of them nonconscious—and are not themselves reducible to experiences. Second, while emotional pleasures tend to be the most important for our hedonic states, they do not exhaust the hedonic realm. Many sensory and “notional” or cognitive pleasures—e.g., approving of a pretty house one passes—seem not in any ordinary sense to be emotional states. Intuitively, they don’t impact our *emotional conditions*. Nor do they seem to affect how happy we are, as illustrated, for instance, by the intense pleasures of sex, which notoriously can be emotionless, leaving us cold even as they feel good. The distinction in question is quite familiar both from common sense and from philosophical and spiritual traditions, where it is widely taken to be an important sign of character what manner of things we allow to get to us, upset us, get us down, raise or lower our spirits, and so forth. Not so for the things that merely give us pleasure or pain, like potato chips and hangnails. This distinction is central to Stoic and Buddhist practices, which focus on the management of our emotional conditions, but expressly let us off the hook regarding mere pleasures and pains. To take a Buddhist example, when pierced by an arrow, the pain is inevitable, but the suffering is optional, and our task is to learn to avoid the emotional disturbance. Aristotle equally took our emotional conditions to be centrally important, though certain sorts of emotional disturbance, like fitting anger, are not to be avoided but actually cultivated. On the emotional state theory, these are all points about the things that make us happy or unhappy; they cannot plausibly be read as claims about pleasure. (Aristotle had a good deal to say about the role of “pleasure” in virtue and hence *eudaimonia* or well-being, but it is clear he meant emotional states; like most commentators, he simply conflates these categories. He certainly is not suggesting that virtuous activity is completed by a pleasing tingle in one’s belly. If the emotional state theory is correct, then what is ordinarily said about the role of pleasure in Aristotelian *eudaimonia* is really about the importance of *happiness* for well-being.)

The foregoing distinction is what I’ve called the *central/peripheral distinction*, where only central affective states—roughly, moods and emotions—constitute our emotional conditions and, in turn, happiness. I argued that central affective states are all “mood-constituting,” so that one’s mood is thereby altered by an emotion. Accordingly, we could just as well speak of mood-related affect. I’ve left it open whether the central/peripheral distinction is merely one of degree, so it is possible that even peripheral affects play some role in happiness and our emotional conditions, though centrality in any event is distinct from degree of pleasantness (e.g., orgasm).

As noted at the outset, some form of emotional state theory has been endorsed by a number of commentators. The variant I have defended, however, adds the idea that happiness is substantially a *dispositional* affair. This figures partly in how I conceive of central affective states, which I take to have dispositional elements that are important for their contribution to how happy one is. When

feeling anxious, for example, one is thereby disposed to respond to events in ways characteristic of anxiety—with more fear, less enjoyment, greater vigilance, and so on. It is partly by virtue of these dispositions that one is less happy when anxious. To this picture I added a further element, *mood propensities*—that is, one’s (appropriately grounded) disposition to experience certain moods and emotions. When depressed, for instance, one may at a given moment be in a good mood, yet still prone to slide back into a flat or depressed mood. One remains depressed—unhappy—by virtue of that propensity alone. In a later paper I allowed that we might substitute the categorical bases of mood propensities for the dispositions themselves—for example, perhaps one is more or less happy by virtue of an unconscious mood state that manifests itself via one’s mood propensity (Haybron, 2010; Hill, 2007). Little hangs on this question, however, so I’ll stick with the original framing.

3. HAPPINESS AS A CONDITION ASSESSMENT CONCEPT

3.1. The problem of dispositionality

We’ve seen that the main source of resistance to the present theory of happiness, even among partisans of an emotional state approach, is the role it accords to dispositional states. Happiness is fundamentally an occurrent phenomenon, the thought goes, with dispositions either being irrelevant or counting only insofar as they are aspects of occurrent emotions and moods.

There is a simple response to such worries: namely, to observe that dispositionality is likewise a central characteristic of the most widely employed conception of happiness in research outside of philosophy, the life satisfaction view. Indeed, life satisfaction may be purely dispositional—essentially a cognitive counterpart to a pure “mood propensity” theory of happiness. It certainly cannot require that any occurrent states be taking place whilst one is happy or unhappy, since the occasions on which we are thinking about how our lives are going will likely be few and far between. To conceive of life satisfaction as an occurrent mental state would yield a theory of happiness with grave “attitude scarcity” problems, whereby most of us are rarely on the happiness-unhappiness scale at all—a result that both is deeply counterintuitive and vitiates the apparent significance of happiness. While life satisfaction theories have come in for strenuous criticism of late, leaving few defenders in the current philosophical literature, the problems have little to do with the dispositionality of life satisfaction.⁸ As far as that goes, the approach is plenty intuitive: *prima facie*, the idea that happiness consists in being satisfied with your life—for instance, in being disposed to judge your life favourably—is quite plausible, to the extent that empirical researchers routinely assume such an account without argument.

While theorizing about the nature of emotional conditions such as depression is less prominent in philosophy, it is very plausible—and indeed is part of the argument for an emotional state view—that depression is substantially dispositional. To be depressed is not merely to feel bad, but for one’s psychological stance

generally to be altered for the worse.⁹ And no one supposes that your depression has resolved if, given the time of day or a particularly agreeable situation, you happen to feel good at the moment. You're still depressed, even when you don't feel it at all. On the emotional state theory, depression just is a pronounced form of unhappiness, and this is not an unintuitive result. Contrary intuitions regarding happiness, then, may owe something to the particular connotations of "happiness," including residual influence from the well-known history of philosophical work using the term in a hedonistic vein. Perhaps it would help to reflect for a bit on life satisfaction views before assessing the emotional state theory.

It is surprising that the inclusion of dispositional states has gotten a mixed reception even among commentators sympathetic to the basic emotional state framework: the basic argument for such a view rests heavily on the failure of hedonism to take seriously the nonconscious aspects of happiness, arguably including dispositionality, in thinking about happiness. If one favours an emotional state theory over hedonism, then hedonism is liable to seem the odd man out, as both emotional state and life satisfaction theories conceive of happiness as a matter of one's orientation or stance toward one's life, whereas hedonism conspicuously does not, reducing happiness instead to a series of purely experiential events.

While dispositionality per se seems to me a positive feature of emotional state theories, and not an objection to them, there seem to me real worries in the neighbourhood. Specifically, the mixed character of the theory, whereby happiness has both occurrent and dispositional aspects, raises questions about how well the various constituents of happiness hang together. Is it just a grue-some assemblage of states rather than a genuine kind? Why not follow the lead of a life satisfaction theory and maintain that happiness is entirely dispositional? While I hope prior argument has made the naturalness of happiness as a kind seem plausible, greater clarity would be helpful.

3.2. Condition assessment concepts

The key lies in a better understanding of what it means to say that happiness is a matter of a person's emotional *condition*, as the emotional state theory was partly defended by adverting to commonsense intuitions about what impacts a person's emotional condition. (The view might thus be more aptly called an "emotional condition" theory of happiness, but that term has unhelpful connotations. It is plausible, at any rate, that a person's emotional state is equivalent to that person's emotional condition.) It would be desirable to have an established theory in metaphysics on the idea of a condition, as opposed, say, to that of an event or state. But as far as I know there isn't one, so I will sketch such an account here. As I mean only to shed light on the nature of a particular sort of condition, happiness, the discussion does not engage very deeply with work in metaphysics and, again, is meant to be only a starting point. But it will, I hope, hold sufficient interest for specialists in that field to warrant further development in the metaphysics literature.

Let's return to the locution "happy and healthy," with which we began. In earlier work I noted that this conjunction signals a likeness of kind between happiness and health. In particular, both appear to involve what we might call *condition assessment concepts* (CACs). Such concepts serve to assess how things stand with functionally organized systems. Are they disposed to function properly? How are they presently configured to function? As noted above, the concept of health seems largely or wholly to concern the individual's disposition: a healthy person is able to function well, is not prone to develop serious problems like a heart attack, and so forth. While the exact contours of the notion of health are much disputed, it is at least strained to regard, say, an episode of indigestion, or soreness or swelling from a stubbed toe as a decline in health. In fact, indigestion or inflammation may indicate *good* health, being healthy responses to the circumstances.¹⁰ These points being noted, I am happy to allow that this notion of health might just be one among others given the diversity of practical concerns that arise in healthcare contexts. A healthy pain response might still merit treatment to ease the patient's suffering.

Used-car buyers and sellers routinely speak explicitly of a car's condition, here employing a CAC appropriate to automobiles: how do things stand with the car? Is it able to serve the purposes for which such vehicles are built? (At least, is it able to serve these purposes relative to the baseline set by nondefective new instances of the model? There's only so much you could expect from a '70s-era Peugeot.¹¹) Can one depend on the engine, the transmission, and other parts to work properly? Cars also serve aesthetic purposes, so the body, paint, and interior should be relatively presentable if it is to be able to serve those ends. Note that not all car problems speak to the car's condition: if the engine is knocking due to bad fuel, that in no way bears on whether the car is in excellent condition; if the engine is generally prone to knocking, that's a reason to downgrade its condition.

With these initial examples in hand, I propose the following definition as a starting point doubtless needing refinement:

functional condition = df that part of the state of a system consisting of the variable dispositional properties that influence its functioning or the processes, parameter settings, or other states that ground those properties

Since "condition" can be used variously, it seems desirable to have a more precise term, hence "functional" condition, though I am not sure the modifier will prove entirely apt, say if there turns out to be a more general metaphysical category at issue. As the definition indicates, functional conditions are a kind of state. They may necessarily be composed of multiple states given the inherent complexity of functionally organized systems, just as a person's health is. This is one reason it is often helpful to employ "emotional condition" rather than "emotional state": the latter is ambiguous, varying between particular emotions, moods, and so forth and a person's overall emotional condition.

Without getting too far into the deep waters of the event/state distinction, we might at least venture this much: conditions are not events; nor are they conjunctions of events. Intuitively, an event is a “happening” and essentially involves temporal structure. When a car engine knocks, that’s an event, a knocking, and these can naturally be counted (“It’s knocking right now, for the second time today”). When a car is in fair condition because its engine tends to knock, that’s not an event, and it isn’t readily dated or counted (“It’s fair-conditioning right now, for the second time today”).

Consider a more pertinent example: pleasures—that is, pleasant experiences—are events, and they can naturally be referred to in the imperfective aspect: “She was experiencing a thrill of pleasure.” (For how long, and when did it start and end?) To be in a certain condition—for instance, to be happy—is not merely for an event to be occurring: “She was happy” need have no temporal structure, such as extension over time or narrative form, and cannot similarly be rendered in the imperfective aspect without loss of meaning.¹² Note that “She was feeling happy” is not at all equivalent, since one can feel happy without being happy, and this is a core feature of the emotional state view. Again, I do not wish to venture too deeply into weeds best navigated by metaphysicians, but it should be plain enough that conditions are not events.¹³

In earlier work I suggested that hedonism about happiness involves a category mistake (Haybron, 2001, 2008b). Even without all the details sorted out, we can now see why: hedonism mistakes a functional condition (the condition of being happy) for a mere conjunction of events (a series of pleasant experiences). Emotional state and life satisfaction theories at least are talking, at some level, about the right sort of thing—roughly, one’s emotional versus cognitive orientation toward one’s life, which essentially includes how one is disposed to respond, emotionally or cognitively, to things.¹⁴ If this is right, then hedonism isn’t merely false; it’s about the wrong kind of thing altogether.¹⁵

3.3. Kinds of functional conditions

The examples of health and cars illustrate one sort of condition: what we might call the *functional soundness* of a system—namely, whether it is disposed to function properly. But many functional systems introduce a further type of condition owing to the fact that they can reconfigure themselves to function differently depending on the circumstances. I will call this, for reasons that will soon be clearer, a system’s *operational status*. Happiness—more generally, a person’s emotional condition—has to do with a person’s operational status, emotionally speaking. That’s not entirely intuitive, so let’s work our way to that idea, starting with simpler examples.

Again, a car. Many vehicles with four-wheel drive can operate in two or more modes: 2- or 4-wheel drive. Depending on which drive mode you’ve selected, the engine will engage the wheels differently. The condition of the system varies depending on the drive mode: it is disposed to function differently in one mode

versus another. Admittedly, the term “condition” is not typically used in this context, though that may reflect the well-entrenched employment of that term to talk about the basic condition of a vehicle—whether it is in good or poor condition, and so forth. As well, it is easy enough to speak directly of the car’s “drive mode.”

The language of conditions becomes more apt when we turn to systems that, like human beings, are equipped with what amounts to evaluative mechanisms, so that the system’s configuration varies with the quality of the conditions it faces: is it configured for good conditions, bad conditions, or...? To illustrate, let’s consider a simplified analogue of a person, a robotic security system whose “emotional” repertoire consists entirely of varying degrees of anxiety (not entirely unlike the robot, Robot, from the television series, *Lost in Space*; so, let’s name it Robot).¹⁶ Its job is to protect a warehouse storing valuable goods from major threats like burglars and minor ones like pests. As it scans the environment over a typical night, it detects the occasional cockroach or other insect or sometimes a misidentified piece of litter, and scoops it up or dispatches it with a laser. A single insect is a trivial matter: an utterly pedestrian negative that gives no reason for concern about the overall situation. It thus merits nothing more than a quick, focused point response to deal with it; the robot does so and resumes its business as before, at its modest baseline level of vigilance (intense monitoring consumes limited battery power and raises the risk of false alarms). Call this a Type-1 response: a transient response to a specific item or event.¹⁷

Suppose Robot detects a break-in. This is a bad situation requiring a broad-based response: sound the alarm! Alert the police, lock down anything that isn’t already secured, and try to frighten off or capture the intruder or intruders. The robot shifts from its baseline low-alert (“relaxed”) to high-alert (“high anxiety”) mode, configured to bring all its resources to bear to protect the warehouse. It scans the environment intently, and small anomalies that might normally be ignored are now treated as threats. Robot remains on high alert for some time, even after police have come and gone. But gradually it shifts to lower levels of readiness until finally resuming its baseline low-alert mode, where it remains until another intrusion is detected. Call this a Type-2 response: a broad-based, sustained response to the quality of the present situation.

There is more to the story. (Let’s assume that pests can be controlled individually and never warrant high alert.) This robot monitors not only the immediate situation, but also the general threat level it is facing these days. If there’s been a rash of intrusions in recent weeks, that suggests that the environment is generally hostile these days. Remaining on elevated alert is costly and increases false alarms, so that isn’t practical. Instead, the robot can adjust the gain on its mechanisms for detecting threats—and/or determining how to respond to apparent threats—so that the mechanisms are more sensitive: it takes less to put the system into high-alert mode. Again, though, this is not an optimal configuration when risks are low, so if things have been quiet for some time, the robot dials down the gain, reducing its propensity to go into high-alert mode, to sound the

alarm, and so forth. We'll call this a Type-3 response: altering the system's preparedness to deal with concerns that may not be live at the moment, but are apt to become so given the general circumstances facing the system.

Security personnel for the warehouse may periodically inquire about the robot's status—in the above terminology, its operational status—to see whether anything is amiss. “What is its condition?” Is it on high alert—“condition red”—indicating a possible break-in? Is it on heightened readiness, suggesting that there may be a higher-than-usual risk of break-ins or a possible vulnerability needing attention?

This framework is quite general and might apply to a wide range of functionally organized systems. Military units, for instance, can instantiate all three types of response, including Type-3 responses, as in the “readiness condition” of a warship. When faced with peacetime conditions, a ship may operate in peacetime cruising condition; in wartime conditions, when hostilities are imminent, the condition may shift to general quarters, with all hands at battle stations, poised to engage in combat operations.

When the health or functional soundness of a system like Robot are not at issue, questions about its “condition” concern not its Type-1 responses, which are merely passing episodes, but its Type-2 and Type-3 responses. Both types of response involve the system's disposition to function, but there is an important difference between them: Type-3 responses are *purely* dispositional, having to do with how the system is disposed to function, while Type-2 responses are also occurrent, involving its functioning at the time as well as its dispositions. It “feels” anxious, so to speak, and this anxiety primes it to respond appropriately to its environment, since it is more prone to assess things negatively, to notice threats, and so forth. It is operating in anxiety mode. As the participle “operating” signals, this aspect of its condition is event-like even if it essentially has dispositional components and so is not merely an event. The system's operational status, then, has two aspects—regarding Type-3 responses, what we may call its *readiness condition* and, regarding Type-2 responses, its *operating mode*.¹⁸ When the robot is sounding the alarm, then, we can say its operating mode is “high alert”; when it is set to high sensitivity to threats, its readiness condition is something along the lines of “primed for high alert.” In some cases, it may be operating in low-alert, relaxed mode, while primed for high alert. Let's call this a “fragile” form of low-alert mode; whereas in “robust” low-alert mode, it is also set for low sensitivity to threats.

I have tried to keep terminology as close to familiar usage as possible, but as we are dealing with a number of distinctions that have not often been made explicit, the jargon is bound to be a chore to navigate. Where feasible, I will refer generically to a system's “functional condition” to simplify the exposition. Before returning to more familiar shores—the case of human happiness—let's briefly summarize the distinctions introduced so far:

Functional condition =_{df} that part of the state of a system consisting of the variable dispositional properties that influence its functioning or the processes, parameter settings, or other states that ground those properties. This can take at least two forms:

1. *Functional soundness*: the disposition of the system to function properly or otherwise
2. *Operational status*: that aspect of a system's condition that determines how it is presently configured to function
 - a. *Readiness condition*: the system's disposition to function in certain modes rather than in others
 - b. *Operating mode*: the mode in which the system is currently operating, which (*inter alia*) disposes it to respond to situations in certain ways

Again, for Robot: its functional soundness amounts to its “health”—what someone looking to buy it would want to know about it; is it in good or poor condition? But when deploying it as a security device, using it partly to inform us about how things are going in its environment, we are typically interested in its operational status, which tells us what alert mode it is in (its Type-2 states or operating mode) or what alert mode it is disposed to assume (its Type-3 states or readiness condition). When it both is on low alert and is disposed to remain on low alert—when it is robustly on low alert—then the robot is, in its sad and uninspiring way, *happy*. Such is the life of Robot.

3.4. Applying the framework to happiness

The terminology is a bit complex, but Robot is not, particularly. No great feat of engineering is required to build such a device, and indeed many examples of similar designs must exist in fact, because this sort of functional organization is an obvious way to build systems that need to respond adaptively depending on the quality of the circumstances they face. It would be very odd were human psychology not able to operate with at least the sophistication of our humble robot.

Even casual observation of human life makes clear that it does. If we return to our primary concern, happiness, the contours of the emotional state theory should already be apparent in the robot example. But it will help to give them greater definition. The distinction between Type-1 and Type-2/3 responses in the robot corresponds to the distinction between “mere pleasures” and happiness:

- Type-1 (point responses): peripheral affective states (mere sensory and notional pleasures and unpleasures)
- Type-2 (operating mode): central affective states (moods and emotions)
- Type-3 (readiness condition): mood propensities

To be angry, for instance, is to be operating in “angry mode,” whereas a different but related question is whether one is poised to switch into angry mode, even when not already angry (say, because one is irritable or prone to irritability), whereas to be in a serene mood is to be operating in “serene mode,” and so forth. To be enjoying a cracker, by contrast, is just to be experiencing pleasure in the eating of a cracker, without this having a direct bearing on your functional condition. To be happy, according to this way of putting things, is to be operating broadly enough in positive versus negative modes and to be disposed to operate that way; it is for one’s operational status, emotionally speaking—that is, one’s emotional condition—to be broadly enough favourable.¹⁹ At least, that is what robust happiness amounts to; one can possess a lesser, fragile kind of happiness when one is configured in a way that favours negative operating modes—that is, moods and emotions—while being lucky enough not to have them triggered. For instance, this might happen during a period of generalized anxiety disorder when one is not, at the moment, anxious, but feels relaxed.

While dispositionality is a hallmark of Type-2/3 states, as well as of happiness, recall that Type-2 states—central affective states—are substantially occurrent states, at least in the typical case. And here we can see the answer to our original puzzle, which was about how happiness could be essentially dispositional, but also partly occurrent: dispositions are often *implemented* by occurrent states. The hypervigilance characteristic of fear, for instance, is part and parcel of the occurrent state of being afraid. The feeling is (or can be) what *makes* you quicker to notice threats. If I annoy my cat, he may shift quite visibly into angry mode; that is, he becomes angry. One more false move and I’m liable to get zapped. Dispositions are essential to this story, but the most salient dispositions can’t enter the picture if nothing is actually happening: the animal must be conscious, operating in some fashion such that he is ready at any moment to zap me if I provoke him further. If happiness were a purely dispositional affair, its effects on our mental states and behaviour would be far remoter and weaker.²⁰

This fact is what gives rise to one of the objections to life satisfaction theories of happiness: as a purely dispositional state—or at least as a state that requires little or nothing occurrent to obtain—it is fairly “causally inert” (Haybron, 2008b). Whether you have a favourable or unfavourable opinion of your life, for instance, may have no bearing on anything if that comes to light only when you are prompted to offer a judgment. Emotions and moods, by contrast, involve a *lot* of internal activity, enabling them to play a profound and pervasive role in regulating one’s other inner states and outward behaviour. Likewise, when Robot sounds the alarm, there is far more going on than merely setting parameters. Robot need not even be functioning for its parameters to be changed, just as you might switch a truck into four-wheel-drive mode when the engine is off. This is why I have claimed that a person can be happy, in a reduced and fragile way, even with a negative mood propensity: *one’s actual moods and emotions do the lion’s share of determining how one is presently disposed to function*. Mood propensity is part of happiness, but a secondary part. Ironically, there is a way in which this view challenges the idea that happiness is a state of passivity:

happiness is not just dispositional, nor just a passive response to one's circumstances. It also consists in occurrent states or processes of emotional functioning, with the organism adaptively reconfiguring itself to better meet the demands of its situation. This emotional activity underwrites various dispositions, but is no less a form of activity for all that.

To see more clearly the distinctions in question, it may help to imagine a simplified state space representation of one dimension of a person's emotional condition: the level of anxiety. The possible states range in a line from zero anxiety to extreme anxiety, and its current level of anxiety is represented by a point on that line—think of it as a rolling ball, resting atop the line—whereas its propensity to varying levels of anxiety is represented by valleys and hills along that line, the valleys being attractors: states toward which the system tends to gravitate; the deeper the valley, the stronger the disposition, and the wider it is, the larger the range of initial states that tend to lead to it. In a case of generalized anxiety disorder, there may be a very wide, deep valley in the region of high anxiety. But at the moment, let's suppose, the person is feeling pretty relaxed, having just gotten a massage. The ball is resting in the low-anxiety region of the line; this is close to the attractor basin (valley) for anxiety, without much of a hill separating it from the attractor, and so this relaxed mood is not a very stable state. But the mood has a bit of stability, and for the time being one is in "relaxed mode," with the according propensities, which we can represent by imagining that the ball has a bit of weight and thus makes a small depression in the line: a little valley. Were this single dimension of anxiety the only aspect of the person's emotional condition, we could say that the individual's happiness is defined by the shape of this state space—the hills and valleys of the line—and the position of the ball. That is, it is defined by the person's mood propensity and current mood.

I have left it open whether peripheral affective states involve the same mechanisms as central affective states, and I have left it open as well whether the central/peripheral distinction is merely one of degree. But for illustrative purposes, let's assume that this anxiety system implements peripheral affects as well—fleeting concerns, say, as when one has a slightly unpleasant experience of noticing an ordinary insect. If the central/peripheral distinction is sharp, then such an experience will register as an effectively weightless dot on the line, with no impact on the line's shape—that is, on one's emotional condition. The dot comes and goes, and that's it, as far as one's emotional state is concerned. But little hangs on whether the central/peripheral distinction is sharp or merely a matter of degree, for in any case the theory requires degrees of centrality, with, for instance, profound anxiety being more central than shallow. In terms of our simplified model, the centrality of an affective state corresponds at least partly to the size of the valley it creates.

In arguing that happiness has this sort of structure, I am of course committing to various empirical claims. In earlier work I noted that my claims about the distinctive functions of different types of affect such as mood do appear to be

consonant with relevant empirical findings.²¹ But the basic framework is highly abstract and theoretical, a conceptual mapping of a sensible and plausible way for many systems, including creatures like us, to be constructed. The conceptual scheme can be implemented in countless ways, in robots, battleships—and, I’m suggesting, sapient apes like ourselves. For the most part the relevant empirical claims—for instance, that our propensities for different moods vary according to the quality of our circumstances—are readily verified from common experience. You don’t need a controlled experiment to figure that out or to know that it will hurt a great deal if you drop an anvil on your foot. There is indeed a great deal of empirical investigation to be done here, if only to determine how exactly these distinctions are implemented in human psychology; the present discussion only circumscribes our subject matter in loose terms. And certain aspects of the view doubtless involve nontrivial predictions that should be testable through scientific methods. But at this stage I think the more pressing questions have to do with the conceptual framework: whether it is coherent, whether it makes the right distinctions, and so forth.

4. HAPPINESS AS A FUNCTIONAL CONDITION: WHAT HAVE WE GAINED?

I hope by this point that the metaphysical status of happiness understood as emotional well-being, in terms of a person’s emotional condition, as well as its difference from happiness as conceived on a hedonistic theory, is reasonably clear. Doubtless more clarity could be brought to the general idea of a functional condition and of its various aspects, as well as to the precise relation between conditions and other states, as well as events. But it makes sense for human affect to be structured along the lines of this emotional state theory, and the dual character of happiness as a hybrid of occurrent and dispositional states should no longer seem ad hoc or chimerical. Certain kinds of functional systems, including human beings, can more adaptively respond to their environments if they are structured in this way, reconfiguring themselves on the fly at multiple levels as conditions demand. It would certainly make no sense for a person to have a fixed trait-level propensity for anxiety, anger, or cheerfulness, no matter what the circumstances. If you’re struggling to get by in war-torn Syria, anxiety should probably come more easily than if you’ve been frolicking for months on a South Seas beach.

But what do we gain by distinguishing the idea of a functional condition or by thinking about happiness in terms of a person’s emotional condition? One benefit, noted earlier, is that happiness, thus understood, enables us to predict and explain many things, such as that a friend is likely to be a pleasanter companion this evening than usual. Our example of Robot illustrates another: such a system’s condition offers an efficient means of gauging how well or badly things are going in the domain the system is meant to evaluate. Security personnel wanting to assess the facility’s threat level might do no better than to check Robot’s condition or status, and indeed that’s an explicit purpose of the system: to summarize and communicate how secure the warehouse is. For human beings,

our emotional conditions do the same, but they add to this many other kinds of information about how well or badly things are going for the individual. In earlier work I suggested that happiness serves as a proxy for well-being: a rough and defeasible indicator of how the person is doing. Rather than just being a security indicator, emotional well-being amounts to an “affective welfarometer” that offers a broad gauge of the individual’s well-being (Haybron, 2008b).²²

It is insufficiently appreciated just how efficient an indicator this sort of condition can be. Suppose our emotional lives were exhausted by the flow of affective experience, so that we only had Type-1 affective responses to our lives. How would you assess how your child or partner is doing these days? You can’t watch them all the time, and you have limited information about their objective situation and, more importantly, the way in which it affects their experience as far as that’s revealed through smiles, frowns, and the like as the day goes on—likely to be an unrepresentative sample as well, since they may tend to feel better (or worse) than usual in your presence. If they get very angry about what seems like a minor event, you can infer nothing other than that they really don’t react well to that sort of thing, or else that their mind was really on something else. It was a bad moment, that’s all. To figure out how they’re doing these days on the basis of this sort of information would require a lot of surveillance—tiring for you, and annoyingly intrusive for them.

But that’s not at all how things work in real life. If your friend blows up over a small matter that appears to be the sort of thing that they normally shrug off with ease, that’s an informative observation: something is amiss. She’s in a bad mood and may thus be having a bad day. Or, if you’ve seen several episodes like this recently, there may be a broader problem: something is probably off in her life these days, and she’s gone into “hostile environment” or “DEFCON 2” mode, as it were, primed to deal with bad events. She’s not happy, and this signals that she’s likely not doing well. More to the point, you were able to make this inference based on very little, easily gleaned information. There is no need for intense scrutiny of her myriad, mostly unobservable feelings over days, nor to compute the integral of such observations. Indeed, only one observation might be needed, say, if she had just burst into tears out of the blue. Similarly, if instead she’d uncharacteristically broken out into song or played a practical joke a few times in the last couple of days, you’d have reason to suppose she’s happy, and that things are going pretty well for her. In essence, an evaluation of how her life is going is encoded in her emotional condition, which in turn disposes her to react to events in certain ways. From such reactions you are able, with little effort and without neuroimaging gear from the future, to make reasonably accurate judgments about her well-being.

“Reasonably accurate”: the claim is not that happiness perfectly tracks well-being.²³ I won’t rehash previous arguments for thinking happiness serves as a proxy for well-being, by, for instance, rebutting dubious “set point” and “happy slave” claims regarding extreme adaptation (Haybron, 2008b). To be sure, there’s no reason to expect our emotional conditions to reflect chronic background

conditions in our lives that should have no bearing on our behaviour or functioning. From a biological standpoint, this would just be a waste of resources. The death of a partner, the failure of a career, or unjust societal conditions might plausibly be thought to affect well-being in lasting ways even when the misfortune has no emotional impact. Such examples help to motivate my own view that happiness is just one aspect of well-being, which also includes value fulfillment: in these cases, important values are being frustrated even if the harm is no longer salient on a daily basis (Haybron, 2022). But even if the loss of a spouse or career leaves a permanent mark, diminishing one's well-being, it seems entirely compatible with doing well on the whole, indeed thriving, just as a disability might diminish well-being without at all precluding flourishing.

The fact that our emotional conditions should not be expected to track closely all aspects of well-being may not be any great concession. If the three-dimensional framework I've suggested for emotional well-being is correct, for instance, then emotional well-being responds to three broad sorts of welfare concerns: security (attunement), opportunity (engagement), and success (endorsement) (Haybron, 2008b). It is not clear that any major domain of well-being, at least of a sort that might be anything like a consensus point, is left out of this framework. Even if emotional well-being doesn't perfectly track well-being in, for instance, ceasing to register long-ago successes and failures, it is not clear that a great deal is left out in most cases.

Here are some external conditions that do seem quite central to well-being, so that if one is doing badly in these areas, one is very likely doing badly, period, however one feels: problems with or relating to relationships, such as a bad marriage or a child in crisis, or problems with one's main occupation in life, be it work or something else. Intuitively, things like a bad marriage or a bad job—or a lack of any meaningful way of passing one's time at all—tend to be incompatible with thriving or well-being. But these are also things that are strongly associated with unhappiness (Haybron, 2013a). And if someone manages to be genuinely happy despite having a bad job, then perhaps the work issue isn't such a big deal—the boss is a jerk, but it's just a job and it pays the bills, and so on. Likewise for a marriage: if the partners are happy, then perhaps they aren't actually doing badly in their lives. (There might well be a specifically *moral* problem, say, if the woman is subservient to the man. While there is no reason to think that people living in oppressive circumstances tend to lead particularly fulfilling lives, there's also little reason to address such cases as being mainly welfare problems.²⁴)

It may be helpful to recall that welfare hedonism is not in fact a nonstarter.²⁵ Like the other major theories, it seems at least to get roughly the right verdict in most cases, which is why people take it seriously. In general, people don't adapt so thoroughly to their life conditions that their enjoyment of life is only loosely connected to how well they're doing. This is both intuitively obvious and amply supported by the data, which reveal vast differences in subjective well-being across different life circumstances.²⁶ The contrary view was only taken seriously

during a brief period a couple of decades ago when a number of researchers went on sort of an intellectual bender, suggesting against all experience with things like marriages and jobs that happiness is largely immutable. Stranger things have happened in the academy, like behaviourism, but that one was up there. Note, by way of illustration, that it took a wild thought experiment involving an experience machine to pose a really serious counterexample to hedonism (Nozick, 1974). If happiness were really only loosely connected to well-being, one shouldn't have to work so hard to come up with counterexamples.

If technology advances sufficiently somehow to allow us to thrive emotionally in lives that are completely decoupled from reality, and if enough people derive their happiness from such sources, then we might have a problem using happiness as a welfare gauge. Until that time it remains plausible that happiness generally tracks well-being well enough to be a rough proxy for it: if someone is happy, they're probably doing well; if unhappy, badly.

While our emotional conditions serve important informational functions, that of course is not their sole or even main purpose, which is broadly to regulate our functioning so that it is appropriate to our goals and circumstances. I believe that this point has ramifications for our understanding of human agency, and that our emotional conditions reflect the character of the self, implementing a person-level, though not quite rational, form of regulation.²⁷ This suggestion requires considerable elaboration, however, and will need to be developed elsewhere.

5. CONCLUSION

5.1 Summary

We began with a puzzle for emotional state theories of happiness, at least of the sort I defend: what is the justification for including dispositional states, including purely dispositional states like mood propensities, in an account of happiness? Heightening the worry is the notion that happiness also includes occurrent states. These questions are all the more pressing when the view is contrasted with an older account, hedonism, that seems traditionally to have been confused with it. The significance of pleasure is precisely that it is *not* dispositional. But, to begin with, this is the wrong contrast: in contemporary thought, both commonsense and academic, the most prominent understanding of happiness has arguably been the life satisfaction theory, which has an entirely different structure from hedonism.²⁸ Indeed, life satisfaction may well be entirely dispositional. To be happy, on that view, is to take a certain stance toward one's life—to be disposed toward it in certain ways.

Once the emotional state theory is distinguished from hedonism, so that we can see how happiness might both be a matter of affect and have dispositional aspects, it becomes evident that hedonism is the outlier: it is odd to call someone

happy simply on the grounds that a pleasant series of experiences has passed through that person's mind lately. That makes happiness too nearly something that happens to a person—a mere agglomeration of experiential events—rather than a genuine state of the person. On reflection, it seems that hedonism isn't even a plausible candidate theory of happiness. As I suggested in earlier work, it is guilty of a category mistake. But the exact nature of that mistake was left unclear: just what categories are being confused?

This paper has ventured an answer: the categories of events and conditions, specifically functional conditions. Whereas hedonism reduces happiness to a series of events, the emotional state theory, like the life satisfaction account, identifies happiness with an aspect of a person's functional condition. As such, the concept of happiness belongs to a family of condition assessment concepts that we use to assess how a functional system is configured to function, with a robotic security system employed to illustrate. While the notion of a functional condition is in great part dispositional, it very often—almost invariably in the case of happiness—consists substantially in occurrent states, as the relevant dispositions are so often implemented by occurrent states. When angry, for instance, you are functioning in “angry mode,” so that the way you feel grounds various dispositions to respond, usually not favourably, to things in your environment. Understood this way, happiness consists both in dispositions and in occurrent functioning.

A great deal is lost if we do not employ—at least tacitly—condition assessment concepts in thinking about well-being, and certainly in thinking about our emotional lives. To assess a person's well-being, we would be left in the hedonist's predicament, attempting to infer from a hopefully representative-enough sample of observed emotions and moods, taken as piecemeal responses to the flow of events in the person's life, how it all adds up. Even if we master this computational task, we still are given little sense of the macro-level picture regarding how the person's life is going—the sort of information one might hope to glean from a life satisfaction judgment, for instance. The emotional information is essentially robbed of context—which might be important in its own right, but is all the more crucial to know insofar as much of the emotional story is not readily observed, say, because the depressed friend usually hides her distress behind an obligatory smile. We are left with a misleadingly shallow and fragmented picture of persons' emotional lives—just one damned thing after another—so that we are bound to underestimate the importance of emotional well-being for human functioning and flourishing.

Finally, I have suggested that our ontology needs supplementation with the category of functional conditions, which among other things allows us more clearly to differentiate an intuitive understanding of a system or person's *state* from that of an *event*, even if in broad terms the event-state distinction is not entirely firm. Given how many of our concerns relate to the way things stand with functionally organized systems, not least ourselves and other people, I expect that the concept of a functional condition will prove fruitful in a variety of contexts.

5.2 Implications for psychological science

I want to close by emphasizing the significance of the gap in psychologists' understanding of the emotional realm, which was noted in earlier work but should now be even more clearly problematic. Contemporary psychology relies on a crude and implausibly sharp distinction between traits and states, where states are wholly occurrent, along the lines of Type-1 responses. Insofar as happiness has dispositional aspects, then, it must be a trait construct. But that is false even to the standard means of operationalizing mental-health constructs like depression and anxiety, which are typically measured with surveys including dispositional items about one's "ability to" laugh, concentrate, sleep, and so forth. Without a category of functional conditions, psychological science lacks the vocabulary even to state what depression, happiness, and other emotional conditions are.

One result of this omission is that emotional well-being measures in well-being research—indulging for the moment the surpassingly weird convention of treating mental-health research as about something other than well-being—invariably do nothing more than inventory the frequency and perhaps intensity of various feelings individuals have experienced during some time frame, as if Type-1 states exhausted the emotional realm. Outside the mental-health paradigm, one seems to find dispositional items only in trait measures, for instance, in personality scales. Yet some such items do not appear to concern traits or personality at all, but rather emotional well-being. The popular International Personality Item Pool (IPIP), for instance, has a suggested fifty-item variant, starting with the unpromising prompt: "Describe yourself as you generally are now, not as you wish to be in the future."²⁹ Even before seeing item one, we can surmise that "as you generally are now" is liable to include your response to present conditions, which may well be unrepresentative (perhaps your home was recently incinerated by a wildfire, and this has gotten your goat for some weeks now). If we take just the first four items, the problem should be apparent enough:

1. Am the life of the party
2. Feel little concern for others
3. Am always prepared
4. Get stressed out easily

It takes little imagination to see how someone who has become uncharacteristically downhearted in the wake of her house burning down might score poorly on these items, though normally she would give the opposite sorts of responses. These items convey good information about one's present emotional disposition and might well constitute part of a good happiness measure. But they tell us little about one's personality.

A good deal of psychological research, in short, is hamstrung by a primitive state/trait distinction that assimilates dispositionality to the realm of traits. It can't even handle a simple and obvious cognitive architecture like Robot's, let

alone a human being's. The ability to reconfigure ourselves according to the conditions is crucial to how we adapt and navigate the world, and a psychology that can't cope with this very basic aspect of human nature is missing quite a lot. Psychology can't do its work without something like the notion of a functional condition. That it has gotten so far without one may owe to the likelihood that you can ignore the dispositional elements of emotional well-being and still get reasonable-looking correlations among self-reports, since one's emotional dispositions will tend to be associated with occurrent emotions, and self-reports of dispositions are not trivial to render and would presumably rely on recalled introspections of occurrent emotions. Likewise, people's recent emotional histories probably correlate fairly well with their personalities; on a typical day, you're liable to feel as you typically do. Moreover, "reasonable-looking correlations" is a pretty elastic notion: in science, lots of findings can seem reasonable, which perhaps makes instruments easier to validate than they should be. So long as one is content with a psychology that merely summarizes correlations among self-reports without attempting to model the underlying processes—as if one were to do cognitive science entirely by word-cloud analysis—the problem may not look so bad.³⁰ But it would be better to understand what's going on under the hood, among other things because this might help us better interpret what people's fallible self-reports, which may uniformly share similar weaknesses, are really telling us.

A more pressing issue is that well-being research may be misconceived so long as it fails to distinguish people's traits and occurrent states from their conditions. Emotional well-being metrics, for instance, are sometimes dismissed even by subjective well-being researchers as concerned merely with momentary feelings that reflect the flux of daily events, but not the global picture regarding the things that matter in life, which is thought to be the domain of life satisfaction metrics. The annual World Happiness Reports, for instance, sideline emotional measures in favour of global life satisfaction—strictly, life evaluation—measures as the chief indicator of well-being. This appears to reflect multiple factors, but at least in conversation the putative transience of emotional states is often mentioned as a major reason for according them secondary status.

In fairness, operationalizing the distinction between occurrent and dispositional emotional states is not trivial. As just noted, emotional-well-being measures may yield similar results in practice with or without that distinction. Asking people about their disposition to feel sad, say, may not get you very different answers in large-scale surveys from what you would find if you simply asked them how often they felt sad.³¹ But this should be noted as a limitation of the measures, and efforts should be made to overcome it. To their credit, mental-health researchers frequently include dispositional items like "found it difficult to relax" in their scales, but the empirical difference made by that format has not, I believe, been very well explored. Other measurement techniques like mood induction—how individuals respond to stressors, say—may also be useful.³² But new measures are not necessarily required: in one time-use study, the effects of unemployment were assessed using common instruments, with life satisfaction measures finding

a plausible reduction in scores among the unemployed; but a hedonic instrument assessing experiences showed little difference between the employed and the unemployed.³³ At the same time, the unemployed derived less pleasure from given activities than the employed. The apparent discrepancy plausibly has a dispositional explanation: unemployed individuals appear to be less happy, with poorer emotional conditions, than their employed counterparts, but compensate by allocating their time to more enjoyable activities, like video gaming rather than sitting in meetings. The example suggests a further avenue for studying the dispositional aspects of emotional well-being, but also illustrates how pleasure and happiness can diverge in ordinary life: the less happy might still have relatively pleasant experiences through judicious time allocations. One could imagine French citizens being less happy than their U.S. counterparts while enjoying a comparable quality of life, because more time is spent on things like sharing meals in France.

Interestingly, what makes emotional conditions informationally useful in everyday life—namely, that we can efficiently assess each other’s emotional responses to our lives from very limited observations—may also pose significant obstacles for the scientific study of happiness. You and I can deduce a great deal about a loved one’s inner emotional life from a single episode, even drawing on information that no one explicitly notices. But such delicate exercises of social-emotional intelligence are no small matter to reproduce in any kind of scientific instrument, let alone a large-scale survey. Be that as it may, it would be reassuring to see some acknowledgement that there is a distinction, as well as some evidence about how best to manage it operationally.

One might ask whether life satisfaction is in fact a better proxy for well-being than emotional well-being, at least in empirical contexts like the World Happiness Reports. Nothing argued here is incompatible with that: there could be more than one useful proxy for well-being, and perhaps the cognitive and affective metrics are complementary. As well, simple measures from vast samples across diverse cultures are needed for these sorts of studies, and perhaps life satisfaction measures would serve better for those purposes. Moreover, problems that might vitiate life satisfaction as a proxy in daily life—for instance, the fact that they are so easily gamed and prone to rationalization and hence of dubious value in the individual case—may tend to wash out over large samples. The epistemic demands of scientific research and everyday human life are quite different. So, it is certainly an open question whether emotional-well-being measures should be treated as a proxy for well-being in data-driven policy contexts, say.

But I do not think we should be too concessive just yet. There is some evidence, for instance, that the affect items used in the Gallup World Poll, which informs the World Happiness Reports, may in fact track well-being better than the life satisfaction measures. For instance Diener and Ng found that the emotion questions were better predicted by “psychosocial prosperity” items, whereas life satisfaction more closely tracked material prosperity—things like income and possessions (Diener, 2010).³⁴ Psychosocial prosperity means things like good

relationships and jobs, noted earlier as among the universally acknowledged nonnegotiables of well-being. The importance of money and stuff for well-being, by contrast, is eminently negotiable, and you might prefer well-being measures that track the obviously important stuff more closely than the dubiously important stuff. More recently, life satisfaction measures astonishingly showed little or no decline in most studies after the onset of the COVID-19 pandemic—possibly a textbook case of adaptive preference-like processes—though emotional-well-being instruments focused on mental health revealed a far grimmer, and more believable, picture of well-being trends, with skyrocketing symptoms of anxiety and depression.³⁵ These of course are just small slices of a very large and complex picture, and again, for current purposes, nothing hangs on whether emotional well-being measures offer the best snapshot of overall well-being. Measurement introduces all sorts of problems that don't concern us here. The more important point is that our picture of human well-being needs a richer understanding of the how a person's emotional life hangs together. It's something deeper than a string of feeling episodes.

NOTES

- ¹ I am grateful to participants at the Conference on Well-Being and Affective States in Clermont-Ferrand, John Doris, and an anonymous referee for comments on an earlier version of this material, and to Ruth Groff, Eric Marcus, John Schwenkler, and Kent Staley for helpful advice on the issues discussed here.
- ² E.g., Davis, 1981; Feldman, 2010; Heathwood, 2022.
- ³ Haybron, 2005, 2008b, 2010. Others endorsing some form of the emotional-state theory appear to include Badhwar, 2014; Becker, 2012; Besser-Jones, 2013; Kauppinen, 2013; Klausen, 2016, 2020; Lazari-Radek & Singer, 2014; May, 2015; Raibley, 2012; Rodogno, Krause-Jensen, & Ashcroft, 2016; Rossi, 2018; Rossi & Tappolet, 2016; Sizer, 2010; Tiberius, 2018; and Višak, 2015. Philosophers who have endorsed alternative conceptions of happiness, typically hedonism or life satisfaction views, since 2008 include Blackson, 2009; Feldman, 2010; Goldman, 2016, 2019; Heathwood, 2022; Morris, 2011; Mulligan, 2016; Skidelsky, 2017; Zamuner, 2013; and perhaps Suikkanen, 2011. Goldman's view is not easily categorized but appears to be a form of life-satisfaction theory. Suikkanen's theory identifies happiness with a hypothetical, idealized life-satisfaction judgment and might most charitably be read as an account of well-being rather than happiness in the psychological sense. Vitrano's life-satisfaction theory seems pretty clearly to use "happiness" in its well-being sense, and so is not a theory of happiness in the present sense (Vitrano, 2010, 2014).
- ⁴ E.g., Heathwood, 2022; Hill, 2007; Klausen, 2016; Morris, 2011; Raibley, 2012; Rossi, 2018.
- ⁵ See note 3 for references, especially Blackson 2009; Feldman, 2010; Morris, 2011.
- ⁶ While "functional condition" has been used, e.g., in discussions of functionalist theories of mind and while the present notion resembles that one, my discussion of functional conditions does not require a functionalist framework.
- ⁷ For reviews of the literature on happiness, see Besser, 2021; Haybron, 2020; Heathwood, 2021. Unless otherwise noted, the characterization of the emotional-state view in this section is based on the discussion in Haybron (2008b).
- ⁸ The classic articulation of a life satisfaction view in the recent philosophical literature is Sumner, 1996; see also references in the previous note. Criticisms include Feldman, 2008, 2010; Haybron, 2007, 2008b, 2011b.
- ⁹ See, e.g., Tully, 2017.
- ¹⁰ In earlier work I suggested that the concept of well-being is a CAC (Haybron, 2008b, p. 142). That may be correct depending on the nature of the concept; we sometimes seem to understand well-being along the lines of a CAC, as assessing a person's overall condition ("How are you?"); but well-being might alternatively—or via a different concept of prudential value—be understood as merely a summary of events, as in hedonic theories, and it may be that the idea of your life going well for you similarly centres on events rather than on one's condition. But in that discussion, I employed a broader understanding of a CAC, as simply an assessment of the practical significance of a person's situation—e.g., as warranting concern or intervention. That now seems a mistake since even summaries of events like hedonic experience can serve that purpose, though they aren't about anyone's condition.
- ¹¹ I'm told they've improved considerably since the 1976 504, which I mention for no particular reason.
- ¹² A condition like being happy, depressed, or healthy will presumably be located in time, and they may well happen to involve some temporal structure while they obtain. But these kinds are not defined in temporal terms, as happenings that are essentially extended in time, with some sort of narrative course (loosely speaking), and so forth. They are in a sense "static" kinds, though this way of putting it may be misleading since again their predication does not preclude that various events are taking place while they obtain.
- ¹³ I am grateful to Eric Marcus for guidance on these matters, though I am not sure he would

agree with my remarks here; for a helpful discussion of the event/state distinction, see Marcus, 2012. It seems to me that whatever might distinguish states from events, the distinction between conditions and events is still clearer.

- ¹⁴ I am simplifying: life satisfaction may have affective and conative aspects as well. While dispositional, life satisfaction might not seem to be plausibly characterized in terms of a person's "condition," especially if it is regarded as little more than a belief. In fact it is a problem for the view that the psychological impact of life satisfaction can be slight (Haybron, 2008b, p. 85). But it need not always be—for instance, when dissatisfaction with one's life motivates one to seek major changes. In that case it is more natural to see life satisfaction as bearing on one's condition. Similarly, beliefs and desires generally bear on one's functional condition insofar as they have dispositional aspects, but in most cases—one's belief that there is no king of France, for instance—the impact is insignificant. In some instances, like acquiring a belief in God's existence, the effects are so far reaching that it is not unintuitive to regard it as a change in condition.
- ¹⁵ Feldman's "attitudinal hedonism" about happiness is not obviously guilty of this charge, but only because it is not clearly an instance of hedonism, and in crucial respects it more closely resembles the life satisfaction view (Feldman, 2010; Haybron, 2008b, p. 65). There is some plausibility, for instance, to the idea that one is happy to the extent that one is pleased with things in one's life. But it is also intuitively plausible that to be pleased with things is at least partly to be disposed toward them in certain ways, so that the concept of attitudinal pleasure may in fact be a CAC.
- ¹⁶ This is an exercise in what Grice called "creature construction" (Bratman, 2000; Grice, 2013; Railton, 2014).
- ¹⁷ "Type-1," etc. as used here should not be confused with the unrelated Type-1/Type-2 distinction from dual-process models of mind.
- ¹⁸ Note that a system's readiness condition could also be sensibly referred to as an operating mode, so the terminology is not as clear as might be desired.
- ¹⁹ Here as before I leave open what exactly the threshold is for counting as happy, though the bar is plainly higher than a bare majority of positive versus negative.
- ²⁰ It bears emphasizing that dispositions can be distinguished at many levels, and it may indeed be possible to distinguish the dispositions constitutive of happiness as relatively "occurrent" dispositions in the sense that they are closely linked to occurrent states and processes. At the other end of the spectrum, highly idealized states such as Suikkanen's idealized form of life satisfaction can also be regarded as dispositions, but of a sort far removed from the actual flow of events: one is disposed to be satisfied, say, if fully informed, reflective, etc.
- ²¹ The literatures on mood and emotion are not easily navigated, as there is no canonical way of conceptualizing the issues. But for some helpful empirical discussions, see, e.g., Eldar, Rutledge, Dolan, & Niv, 2015; Fox, Lapate, Shackman, & Davidson, 2018; W. N. Morris, 1999; Nettle & Bateson, 2012; Parkinson, Totterdell, Briner, & Reynolds, 1996; Siemer, 2009, and especially Robinson, 2000, which argues specifically that moods serve as summary indicators of recent experience. I discuss this evidence in Haybron, 2014, as well as in Haybron, 2008b. Relevant philosophical discussions include, e.g., Delancey, 2006; Kurth, 2018; Lormand, 1985; Rossi, 2021; Sizer, 2000; and Wong, 2016.
- ²² At least, relative to conventional views of well-being, as opposed to, say, Stoicism.
- ²³ Cf. Raibley, 2012.
- ²⁴ Even consequentialists will standardly grant that one's opposition to injustices like racism should not depend on utility calculations (Railton, 1984).
- ²⁵ Despite the crucial differences between emotional well-being and pleasure, the two goods are plainly strongly correlated, so that if one is a proxy for well-being, the other probably is as well. In fact, probably all major theories of well-being tend in practice to centre on goods that are strongly correlated, thus tending to agree about the majority of cases; for the most part practical disagreements tend to arise at the margins.
- ²⁶ For reviews of empirical findings on various aspects of well-being, see Diener, Oishi, & Tay,

2018. The annual World Happiness Reports from 2012 onward find large differences globally on a variety of well-being indicators (Helliwell, Layard, & Sachs, 2012). While I have noted various limitations of the measures in other work, they do not materially affect the point I'm making here.

²⁷ On the connection between happiness and the self, see Haybron, 2008b, 2008a, 2022. I take the suggested picture of agency to be roughly complementary to John Doris's (Doris, 2002, 2015, 2021), as well as my earlier work on the limitations of rational control (Haybron, 2008b, 2011a, 2014).

²⁸ Though there is some evidence from studies of folk views of happiness that an emotional state conception is actually predominant in lay usage (Kneer & Haybron, 2019).

²⁹ E.g., https://ipip.ori.org/new_ipip-50-item-scale.htm. Accessed June 2, 2021.

³⁰ Alexandrova, 2017; Alexandrova & Haybron, 2016.

³¹ My own work with David Yaden on an emotional-well-being measure suggests this may in fact be the case (Yaden & Haybron, 2022). Only one of the dispositional items we tested ("Been able to laugh about lots of things") was sufficiently predictive to remain in the scale, and the resulting measure, along with other standard affect measures used in subjective well-being research like PANAS and SPANE, correlated very strongly with standard depression and anxiety scales.

³² See, for instance, Shedler, Mayman, & Manis, 1993; Weinberger, Schwartz, & Davidson, 1979.

³³ Knabe, Rätzel, Schöb, & Weimann, 2010.

³⁴ Other studies drawing on Gallup World Poll data found similar results, as did my studies with Yaden, not fully reported in our paper (Ng & Diener, 2014; Tay & Diener, 2011; Yaden & Haybron, 2022).

³⁵ E.g., Helliwell, Layard, Sachs, & De Neve, 2021; Panchal, Kamal, Cox, & Garfield, 2021.

REFERENCES

- Alexandrova, Anna, *A Philosophy for the Science of Well-Being*, New York, Oxford University Press, 2017.
- Alexandrova, Anna and Daniel M. Haybron, "Is Construct Validation Valid?" *Philosophy of Science*, vol. 83, no. 5, 2016, p. 1098-1109.
- Badhwar, Neera K., *Well-Being: Happiness in a Worthwhile Life*, New York, Oxford University Press, 2014.
- Becker, Lawrence C., *Habilitation, Health, and Agency*, New York, Oxford University Press, 2012.
- Besser, Lorraine L., *The Philosophy of Happiness: An Interdisciplinary Introduction*, New York, Routledge, 2021.
- Besser-Jones, Lorraine, "The Pursuit and Nature of Happiness", *Philosophical Topics*, vol. 41, no. 1, 2013, p. 103-121.
- Blackson, Thomas, "On Feldman's Theory of Happiness", *Utilitas*, vol. 21, no. 03, 2009, p. 393-400.
- Bratman, Michael E., "Valuing and the Will", *Philosophical Perspectives*, vol. 14, 2000, p. 249-265.
- Davis, Wayne, "A Theory of Happiness", *American Philosophical Quarterly*, vol. 18, no. 2, 1981, p. 111-120.
- Delancey, Craig, "Basic Moods", *Philosophical Psychology*, vol. 19, no. 4, 2006, p. 527-538.
- Diener, Ed, Weiting Ng, James Harter and Raksha Arora, "Wealth and Happiness across the World: Material Prosperity Predicts Life Evaluation, whereas Psychosocial Prosperity Predicts Positive Feeling", *Journal of Personality and Social Psychology*, vol. 99, no. 1, 2010, p. 52-61.
- Diener, Ed, Shigehiro Oishi, and Louis Tay (eds.), *Handbook of Well-Being*, Salt Lake City, UT, DEF Publishers, 2018. Retrieved from <http://nobascholar.com/books/1>
- Doris, John M., *Lack of Character*, New York, Cambridge University Press, 2002.
- , *Talking to Our Selves*, New York, Oxford University Press, 2015.
- , *Character Trouble: Undisciplined Essays on Moral Agency and Personality*, New York, Oxford University Press, 2021.
- Eldar, Eran, Robb B. Rutledge, Raymond J. Dolan and Yael Niv, "Mood as Representation of Momentum", *Trends in Cognitive Sciences*, vol. 20, no. 1, 2015, p. 15-24.
- Feldman, Fred, "Whole Life Satisfaction Concepts of Happiness", *Theoria*, vol. 74, no. 3, 2008, p. 219-238.

———, *What Is This Thing Called Happiness?*, New York, Oxford University Press, 2010.

Fox, Andrew S., Regina C. Lapate, Alexander J. Shackman and Richard J. Davidson (eds.), *The Nature of Emotion: Fundamental Questions* (2nd edition), New York, Oxford University Press, 2018.

Goldman, Alan H. “Happiness is an Emotion”, *The Journal of Ethics*, vol. 21, no. 1, 2017, 1–16.

———, *Life's Values: Pleasure, Happiness, Well-Being, and Meaning*, New York, Oxford University Press, 2019.

Grice, Paul, “Method in Philosophical Psychology (From the Banal to the Bizarre)” *The American Philosophical Association Centennial Series*, 2013, p. 305-332.

Haybron, Daniel M., “Happiness and Pleasure”, *Philosophy and Phenomenological Research*, vol. 62, no. 3, 2001, p. 501-528.

———, “On Being Happy or Unhappy”, *Philosophy and Phenomenological Research*, vol. 71, no. 2, 2005, p. 287-317.

———, “Life Satisfaction, Ethical reflection, and the Science of Happiness”, *Journal of Happiness Studies*, vol. 8, no. 1, 2007, p. 99-138.

———, “Happiness, the Self and Human Flourishing”, *Utilitas*, vol. 20, no. 1, 2008a, p. 21-49.

———, *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, New York, Oxford University Press, 2008b.

———, “Mood Propensity as a Constituent of Happiness: A Rejoinder to Hill”, *Journal of Happiness Studies*, vol. 11, no. 1, 2010, p. 19-31.

———, “Central Park: Nature, Context, and Human Wellbeing” *International Journal of Well-being*, vol. 1, no. 2, 2011a, p. 235-254.

———, “Taking the Satisfaction (and the Life) out of Life Satisfaction” *Philosophical Explorations*, vol. 14, no. 3, 2011b, p. 249-262.

———, *Happiness: A Very Short Introduction*, New York, Oxford University Press, 2013a.

———, “The Value of Positive Emotion: Philosophical Doubts and Reassurances”, in June Gruber and Judith Tedlie Moskowitz (eds.), *Positive Emotion: Integrating the Light Sides and Dark Sides*, New York, Oxford University Press, 2014, p. 281–300.

———, “Adventures in Assisted Living: Well-Being and Situationist Psychology”, in Nancy E. Snow and Franco V. Trivigno (eds.), *The Philosophy and Psychology of Character and Happiness*, New York, Routledge, 2014, p. 1-25.

———, “Happiness”, in Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, 2020.

———, “Well-Being: Taking Our Selves Seriously”, in John J. Stuhr (ed.), *Philosophy and Human Flourishing*, New York, Oxford University Press, 2022, p. 90-117.

Heathwood, Chris, *Happiness and Well-Being*, New York, Cambridge University Press, 2021.

———, “Happiness and Desire Satisfaction”, *Noûs*, vol. 56, no. 1, 2022, p. 57-83.

Helliwell, John, Richard Layard, and Jeffrey Sachs (eds), *World Happiness Report 2012*, New York, UN Sustainable Development Solutions Network, 2012.

Helliwell, John, Richard Layard, Jeffrey Sachs, and Jan-Emmanuel De Neve, (eds.), *World Happiness Report 2021*. New York: Sustainable Development Solutions Network, 2021.

Hill, Scott, “Haybron on Mood Propensity and Happiness” *Journal of Happiness Studies*, vol. 10, no. 2, 2009, p. 215-228.

Kauppinen, Antti, “Meaning and Happiness”, *Philosophical Topics*, vol. 41, no. 1, 2013, p. 161-185.

Klausen, Søren Harnow, “Happiness, Dispositions and the Self”, *Journal of Happiness Studies*, vol. 17, no. 3, 2016, p. 995-1013.

———, “Understanding Older Adults’ Wellbeing from a Philosophical Perspective” *Journal of Happiness Studies*, vol. 21, 2020, p. 2629-2648.

Knabe, Andreas, Steffen Rätzel, Ronnie Schöb, and Joachim Weimann, “Dissatisfied with Life but Having a Good Day: Time-Use and Well-Being of the Unemployed”. *The Economic Journal*, vol. 120, no. 547, 2010, p. 867-889.

Kneer, Markus, and Daniel M. Haybron, “Happiness and Well-Being: Is It All in Your Head? Evidence from the Folk”, manuscript, 2019. <https://doi.org/10.13140/RG.2.2.12958.69448>

Kurth, Charlie. *The Anxious Mind: An Investigation into the Varieties and Virtues of Anxiety*, Cambridge, MA, MIT Press, 2018.

Lazari-Radek, Katarzyna de, and Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics*, New York, Oxford University Press, 2014.

Lormand, Eric, “Toward a Theory of Moods”, *Philosophical Studies*, vol. 47, no. 3, 1985, p. 385-407.

Marcus, Eric, *Rational Causation*. Cambridge, MA, Harvard University Press, 2012.

May, Todd, *A Significant Life*, Chicago, University of Chicago Press, 2015.

Morris, Stephen, “In Defense of the Hedonistic Account of Happiness”, *Philosophical Psychology*, vol. 24, no. 2, 2011, p. 261-281.

Morris, William N., “The Mood System”, in Daniel Kahneman, Ed Diener and Norbert Schwarz (eds.), *Well-Being: Foundations of Hedonic Psychology*, New York, Russell Sage Foundation, 1999, p. 169-189.

Mulligan, Kevin, “Happiness, Luck and Satisfaction”, *Argumenta*, vol. 1, no. 2, 2016, p. 133-145.

Nettle, Daniel and Melissa Bateson, "The Evolutionary Origins of Mood Review and Its Disorders", *Current Biology*, vol. 22, no. 17, 2012, p. R712–R721.

Ng, Weiting and Ed Diener, "What Matters to the Rich and the Poor? Subjective Well-Being, Financial Satisfaction, and Postmaterialist Needs across the World", *Journal of Personality and Social Psychology*, vol. 107, no. 2, 2014, p. 326–338.

Nozick, Robert, *Anarchy, State, and Utopia*. New York, Basic Books, 1974.

Panchal, Nirmita, Rabah Kamal, Cynthia Cox, and Rachel Garfield, "The Implications of COVID-19 for Mental Health and Substance Use." *Kaiser Family Foundation*, vol. 21, 2021. Retrieved from <https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/>

Parkinson, Brian, Peter Totterdell, Rob B. Briner, and Shirley Reynolds, *Changing Moods: The Psychology of Mood and Mood Regulation*, New York, Addison Wesley, 1996.

Raibley, Jason R., "Happiness Is Not Well-Being" *Journal of Happiness Studies*, vol. 13, no. 6, 2012, p. 1105–1129.

Railton, Peter, "Alienation, Consequentialism, and the Demands of Morality" *Philosophy & Public Affairs*, vol. 13, no. 2, 1984, p. 134–171.

———, "Reliance, Trust, and Belief", *Inquiry*, vol. 57, no. 1, 2014, p. 122–150.

Robinson, Michael D. "The Reactive and Prospective Functions of Mood: Its Role in Linking Daily Experiences and Cognitive Well-Being" *Cognition & Emotion*, vol. 14, no. 2, 2000, p. 145–176.

Rodogno, Raffaele, Katrine Krause-Jensen, and Richard E. Ashcroft, "'Autism and the Good Life': A New Approach to the Study of Well-Being", *Journal of Medical Ethics*, vol. 42, no. 6, 2016, p. 401–408.

Rossi, Mauro, "Happiness, Pleasures, and Emotions" *Philosophical Psychology*, vol. 31, no. 6, 2018, p. 898–919.

———, "A Perceptual Theory of Moods", *Synthese*, vol. 198, 2021, p. 7119–7147.

Rossi, Mauro and Christine Tappolet, "Virtue, Happiness, and Well-Being" *The Monist*, vol. 99, no. 2, 2016, p. 112–127.

Shedler, Jonathan, Martin Mayman, and Melvin Manis, "The Illusion of Mental Health", *American Psychologist*, vol. 48, no. 11, 1993, p. 1117–1131.

Siemer, Matthias, "Mood Experience: Implications of a Dispositional Theory of Moods", *Emotion Review*, vol. 1, no. 3, 2009, p. 256–263.

Sizer, Laura, "Towards a Computational Theory of Mood" *The British Journal for the Philosophy of Science*, vol. 51, no. 4, 2000, p. 743–770.

———, "Good and Good For You: An Affect Theory of Happiness" *Philosophy and Phenomenological Research*, vol. 80, no. 1, 2010, p. 133–163.

Skidelsky, Edward, "Happiness, Pleasure, and Belief", *Australasian Journal of Philosophy*, vol. 95, no. 3, 2017, p. 1-13.

Suikkanen, Jussi, "An Improved Whole Life Satisfaction Theory of Happiness", *International Journal of Wellbeing*, vol. 1, no. 1, 2011, p. 1-18.

Sumner, L. W. *Welfare, Happiness, and Ethics*. Oxford, Clarendon Press, 1996.

Tay, Louis and Ed Diener, "Needs and Subjective Well-Being around the World", *Journal of Personality and Social Psychology*, vol. 101, no. 2, 2011, p. 354-365.

Tiberius, Valerie, *Well-Being as Value Fulfillment*, New York, Oxford University Press, 2018.

Tully, Ian, "Depression and the Problem of Absent Desires" *Journal of Ethics and Social Philosophy*, vol. 11, no. 2, 2017, p. 1-16.

Višak, Tatjana, "Does Welfare Trump Freedom? A Normative Evaluation of Contextualism about How to Promote Welfare", in Gottfried Schweiger and Gunter Graf (eds.), *The Well-Being of Children*, Warsaw, De Gruyter Open, 2015, p. 34-48.

Vitrano, Christine, "The Subjectivity of Happiness", *The Journal of Value Inquiry*, vol. 44, no. 1, 2010, p. 47-54.

———, *The Nature and Value of Happiness*, Boulder, Westview Press, 2014.

Weinberger, Daniel A., Gary E. Schwartz, and Richard J. Davidson, "Low-Anxious, High-Anxious, and Repressive Coping Styles: Psychometric Patterns and Behavioral and Physiological Responses to Stress", *Journal of Abnormal Psychology*, vol. 88, no. 4, 1979, p. 369-380.

Wong, Muk Yan, "The Mood-Emotion Loop", *Philosophical Studies*, vol. 173, no. 11, 2016, p. 3061-3080.

Yaden, David B. and Daniel M. Haybron, "The Emotional State Assessment Tool: A Brief, Philosophically Informed, and Cross-Culturally Sensitive Measure", *The Journal of Positive Psychology*, vol. 17, no. 2, 2022, p. 151-165.

Zamuner, Edoardo, "Happiness, Consciousness, and the Ontology of Mind", *Philosophical Topics*, vol. 41, no. 1, 2013, p. 237-254.

UNE THÉORIE DU BIEN-ÊTRE COMME BONHEUR APPROPRIÉ¹

MAURO ROSSI

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

CHRISTINE TAPPOLET

UNIVERSITÉ DE MONTRÉAL

RÉSUMÉ :

Il existe un lien intuitif entre le bien-être et le bonheur. Nous souhaitons proposer une nouvelle théorie selon laquelle le bien-être consiste en un *bonheur approprié*. Notre théorie peut être considérée comme la combinaison de quatre thèses. La première thèse est que le bonheur psychologique consiste en une balance largement positive d'états affectifs tels que les émotions, les humeurs et les plaisirs sensoriels. La seconde est que les émotions, les humeurs et les plaisirs sensoriels sont différents types d'expériences perceptuelles de propriétés évaluatives. Il s'ensuit que le bonheur consiste en une balance largement positive d'expériences perceptuelles de propriétés évaluatives. La troisième thèse est que, dans la mesure où le bonheur est constitué par des états qui ont des conditions d'adéquation, alors le bonheur aussi peut être évalué comme étant approprié ou inapproprié. La quatrième thèse est une thèse concernant le bien-être : elle affirme que le bien-être consiste en un bonheur approprié ainsi défini.

ABSTRACT:

There is an intuitive connection between well-being and happiness. Our goal is to defend a new theory, according to which well-being consists in fitting happiness. Our theory can be considered as the combination of four claims. The first is that happiness consists in a broadly positive balance of affective states such as emotions, moods, and sensory pleasures. The second is that emotions, moods, and sensory pleasures are different kinds of perceptual experiences of evaluative properties. It follows from this that happiness consists in a broadly positive balance of perceptual experiences of evaluative properties. The third is that insofar as happiness is constituted by states that have fittingness conditions, happiness can be assessed as fitting or unfitting. The fourth claim concerns well-being: we claim that well-being consists in fitting happiness as we have defined it.

INTRODUCTION

Il existe un lien intuitif entre le bien-être et le bonheur. À tel point que ces termes sont souvent pris pour des synonymes. Ce n'est pas ainsi que nous les utiliserons ici. Dans cet article, nous considérerons le terme « bien-être » comme un terme évaluatif, faisant référence à un type spécifique de valeur, à savoir la valeur prudentielle, tandis que nous considérerons « bonheur » comme un terme psychologique, faisant référence à un état mental favorable ou à une combinaison d'états mentaux favorables². Ainsi conçue, la relation entre le bien-être et le bonheur consiste, si elle existe, en une relation substantielle. De nombreux théoriciens du bien-être ont affirmé l'existence d'une telle relation. On trouve dans la littérature deux grands types de théories du bien-être fondées sur le bonheur. La première soutient que le bien-être n'est rien d'autre que le bonheur tout court. On les appelle les « théories standard du bien-être basées sur le bonheur ». La seconde considère que le bien-être consiste en un bonheur adéquatement qualifié, par exemple un bonheur authentique, un bonheur mérité, etc. Ces théories du bien-être sont appelées « théories hybrides du bien-être basées sur le bonheur ». Les théories standard ont été attaquées parce qu'elles sont soit inadéquates sur le plan descriptif, dans le sens qu'elles ne correspondent pas à nos jugements paradigmatiques sur le bien-être, soit inadéquates sur le plan normatif, dans le sens qu'elles ne sont pas utiles d'un point de vue normatif. Ces objections ont motivé l'élaboration de théories hybrides. Pourtant, ces dernières ont également été contestées. La principale objection soulevée à leur endroit est qu'elles sont *ad hoc*, c'est-à-dire qu'elles combinent des éléments de différentes théories du bien-être d'une manière qui n'est pas justifiée de façon indépendante. Cet état de fait soulève la question de savoir s'il existe une théorie fondée sur le bonheur qui puisse être défendue avec succès.

Dans cet article, nous souhaitons faire un premier pas vers une réponse positive à cette question. Plus précisément, nous souhaitons proposer une nouvelle théorie selon laquelle le bien-être consiste en *un bonheur approprié (ou adéquat, correct)*³. Notre objectif dans cet article est de présenter les principales caractéristiques de cette théorie. Les étapes qui restent à franchir consistent à montrer que notre théorie est immunisée contre toutes les objections soulevées à l'encontre des théories traditionnelles basées sur le bonheur et, finalement, qu'elle est préférable à toutes les autres théories du bien-être. Ces tâches dépassant nettement le cadre d'un seul article, nous les réservons donc à d'autres travaux⁴.

Notre théorie peut être considérée comme la combinaison de quatre thèses, que nous présenterons séparément dans les sections suivantes. La première thèse est que le bonheur psychologique consiste en une balance largement positive d'états affectifs tels que les émotions, les humeurs et les plaisirs sensoriels (section 1). La seconde est que les émotions, les humeurs et les plaisirs sensoriels sont différents types d'expériences perceptuelles de propriétés évaluatives (section 2). Il s'ensuit que le bonheur consiste en une balance largement positive d'expériences perceptuelles de propriétés évaluatives. La troisième thèse est que, dans la mesure où le bonheur est constitué par des états qui ont des conditions d'adé-

quation (*fittingness conditions*), alors le bonheur aussi peut être évalué comme étant approprié ou inapproprié (*fitting or unfitting*). En particulier, nous soutenons que le bonheur est approprié dans la mesure où ses constituants sont appropriés (section 3). La quatrième thèse est une thèse concernant le bien-être : elle affirme que le bien-être consiste en un bonheur approprié ainsi défini (section 4). Nous concluons en énumérant une série d'objections soulevées à l'encontre des théories traditionnelles fondées sur le bonheur et auxquelles nous devons répondre pour défendre avec succès notre théorie.

1. LE BONHEUR PSYCHOLOGIQUE

Notre théorie du bonheur appartient à la famille des théories émotionnelles du bonheur, dont la plus populaire est la théorie proposée par Daniel Haybron (2008). Afin de motiver notre théorie, il est donc utile de la comparer à celle de Haybron. Selon Haybron, le bonheur consiste en une balance largement positive d'états affectifs occurrents, tels que les émotions et les humeurs, et d'états purement dispositionnels, tels que les propensions à ressentir des humeurs. Pour commencer, clarifions la terminologie : les états occurrents sont des épisodes mentaux qui possèdent une phénoménologie spécifique ; en revanche, les états purement dispositionnels sont des dispositions à ressentir des états occurrents, dépourvus de phénoménologie en dehors de celle de leurs manifestations. Selon Haybron, le bonheur est déterminé à la fois par l'intensité et par la centralité de ces états constitutifs du bonheur. Par intensité, il entend l'intensité phénoménale, c'est-à-dire la mesure dans laquelle un état est « ressenti » d'une certaine manière. Par centralité, Haybron entend principalement la disposition à causer d'autres états affectifs ou des actions particulières.

La théorie de Haybron vise à offrir une conception du bonheur qui, en plus de correspondre à nos intuitions paradigmatiques, peut rendre justice au rôle que le bonheur joue dans l'explication du comportement des gens et dans leur délibération pratique, et qui peut être utile pour la théorisation normative. En faisant de la centralité une caractéristique déterminante du bonheur, Haybron est capable de montrer comment nous pouvons prédire et expliquer le comportement en faisant appel à des attributions de bonheur, et comment le bonheur peut fonctionner comme une considération importante dans la prise de décision et la théorisation normative.

Nous sommes d'accord avec l'approche affective de Haybron et avec l'accent qu'il met sur la notion de centralité. Cependant, nous pensons que la théorie de Haybron est confrontée à certains problèmes, dont nous avons discuté ailleurs (voir Rossi, 2018 ; Rossi et Tappolet, Manuscrit-a). Pour l'instant, considérons simplement ce qui suit. Selon la théorie de Haybron, les états affectifs, tant occurrents que dispositionnels, sont des déterminants du bonheur à part entière. Il en découle que cette théorie classe comme très heureux un individu qui a de fortes propensions aux humeurs positives, mais qui, pour certaines raisons (par exemple parce qu'il est endormi, dans le coma, ou simplement dans des circonstances défavorables qui empêchent lesdites propensions de se manifester), n'a

aucun état affectif occurrent positif, c'est-à-dire aucune émotion ou humeur positive. Ce résultat est contre-intuitif. En effet, le bonheur semble avoir une dimension phénoménale essentielle. La leçon à tirer est donc la suivante : nous avons besoin d'une théorie qui préserve l'intuition de Haybron sur l'importance de la centralité dans le bonheur, mais qui puisse également rendre justice à la phénoménologie du bonheur. C'est ici que notre théorie entre en jeu.

Selon notre théorie, le bonheur consiste en une balance largement positive d'états affectifs occurrents tels que les émotions, les humeurs et les plaisirs sensoriels. Selon nous, les émotions et les humeurs positives, ainsi que les plaisirs sensoriels, sont constitutifs du bonheur ; les émotions et les humeurs négatives, ainsi que les déplaisirs sensoriels, sont constitutifs du malheur. (Par souci de simplicité, dans ce qui suit nous nous concentrerons uniquement sur le bonheur.) Nous sommes d'accord avec Haybron sur le fait que l'intensité et la centralité sont deux dimensions importantes des états constitutifs du bonheur et qu'elles sont pertinentes pour déterminer le degré de bonheur d'un individu. Cependant, nous caractérisons la notion de centralité différemment de Haybron.

Considérez ceci. En général, lorsque nous désirons être heureux, nous ne souhaitons pas faire l'expérience d'une série d'états positifs déconnectés et éphémères. Nous désirons plutôt faire l'expérience d'états positifs qui ont des bases solides et une sorte de cohérence interne. Quand les états constitutifs du bonheur possèdent-ils ces propriétés ? Selon nous, c'est le cas lorsque les états constitutifs du bonheur sont fondés sur les valeurs de l'individu et les reflètent. Pour plus de clarté, nous proposons de distinguer entre *ce que l'individu valorise* et *l'attitude de l'individu qui valorise* ces choses. Ce qui importe pour notre propos, ce sont les attitudes de l'individu qui valorise quelque chose. Les valorisations sont des états mentaux stables, qui jouent un rôle important dans la délibération pratique et qui sont étroitement liés à l'identité d'une personne. Les valorisations ont également des propriétés dispositionnelles robustes, car elles disposent l'individu à expérimenter une variété d'états affectifs liés aux objets valorisés. En outre, elles fournissent une sorte d'« unité thématique » à ces états⁵. À partir de là, nous pouvons reformuler notre affirmation précédente en disant que, lorsque nous désirons le bonheur, nous désirons généralement faire l'expérience d'états qui sont basés sur et qui reflètent nos valorisations. Nous sommes d'autant plus heureux que notre bonheur est ainsi fondé.

Cela étant dit, nous pouvons distinguer deux sens pertinents de la notion de centralité. D'une part, il y a ce que nous appellerons la *centralité de sortie*. Il s'agit de la mesure dans laquelle un état est causalement productif d'autres états et comportements. Cela correspond à la conception de la centralité de Haybron. D'autre part, il existe ce que nous appellerons la *centralité de source*. Il s'agit de la mesure dans laquelle un état est basé sur d'autres états centraux de sortie. Les valorisations peuvent être considérées comme les états les plus centraux en termes de sortie. Ce sont en effet les attitudes qui guident et motivent le plus fondamentalement notre comportement et qui ainsi définissent notre identité. Quant aux états constitutifs du bonheur, ils peuvent être soit centraux de sortie,

soit centraux de source, soit les deux. Par conséquent, si ce que nous avons dit ci-dessus est correct, ce qui importe pour le bonheur, c'est principalement de savoir si, et dans quelle mesure, ces états sont basés sur les valorisations de l'individu et les reflètent. En d'autres termes, ce qui importe pour le bonheur, c'est de savoir si leurs états constitutifs sont centraux de source. Selon nous, le bonheur est donc fonction de l'intensité phénoménale et de la centralité de source des états constitutifs du bonheur.

Cette idée distingue notre théorie de celle de Haybron. Nous pensons que notre théorie est parfaitement capable de rendre justice au rôle que le bonheur joue dans l'explication du comportement des gens et dans la délibération pratique. Dans la mesure où les états constitutifs du bonheur sont basés sur les états dispositionnellement robustes que sont les valorisations, nous pouvons effectivement formuler une série de prédictions sur la façon dont un individu agira, de même que des explications sur son comportement passé, à partir d'attributions de bonheur individuel. De même, nous pouvons expliquer l'importance du bonheur pour la délibération pratique en nous référant au lien entre les états constitutifs du bonheur et les valorisations de l'individu. En général, plus le bonheur est basé sur ces derniers états, plus il sera important pour l'individu qui délibère. En outre, contrairement à la théorie de Haybron, notre théorie préserve l'intuition que le bonheur a une dimension phénoménale inhérente, puisque nous maintenons que le bonheur est constitué d'états affectifs occurrents. Ce faisant, elle évite les implications contre-intuitives de la théorie de Haybron.

2. LA NATURE DES ÉTATS CONSTITUTIFS DU BONHEUR

Passons à notre thèse suivante, concernant la nature des états constitutifs du bonheur. Il s'agit de la thèse la plus distinctive qui sous-tend notre théorie. Nous soutenons que tous les états constitutifs du bonheur, c'est-à-dire les émotions, les humeurs et les plaisirs sensoriels, sont des sortes d'évaluations affectives, c'est-à-dire des évaluations ayant une dimension phénoménale. Prenons d'abord les émotions. Considérez un épisode émotionnel tel que votre peur d'être attaqué par un chien. Le chien qui vous attaque est l'objet intentionnel de votre peur, dans le sens où votre peur concerne (c'est-à-dire est dirigée vers) le chien qui vous attaque. Mais votre peur n'est pas une attitude simplement neutre envers le chien. Au contraire, votre peur est un moyen de vous présenter le chien comme effrayant. Être effrayant est une propriété évaluative. Ainsi, en présentant le chien comme effrayant, votre peur vous fournit une évaluation du chien. De plus, elle fournit une évaluation qui est « ressentie » de manière caractéristique, puisque la peur s'accompagne généralement d'une phénoménologie qui lui est propre. On pense généralement que la propriété d'être effrayant est ce qui permet d'identifier tous les épisodes de peur comme appartenant à un seul et même type d'émotion. Tous ces épisodes sont des exemples de peur parce qu'ils présentent leurs objets intentionnels comme étant effrayants. Ainsi, l'effrayant est généralement appelé « l'objet formel » de la peur. La notion d'objet formel est souvent considérée comme étant utile pour distinguer les différents types d'émotions⁶. Ainsi, tous les épisodes de joie sont des cas du même type d'émotion, à savoir

la joie, en vertu du fait qu'ils présentent leurs objets intentionnels comme ayant la même propriété évaluative, à savoir le joyeux. Tous les épisodes de colère sont des cas du même type d'émotion, c'est-à-dire la colère, en vertu du fait qu'ils présentent leurs objets intentionnels comme étant offensants. Et ainsi de suite. En généralisant, nous pouvons dire que les émotions consistent en des représentations phénoménologiquement saillantes de leurs objets intentionnels comme possédant différentes propriétés évaluatives. En tant que tels, différents types d'émotions comptent comme différents types d'évaluations affectives.

L'idée selon laquelle les émotions sont des évaluations affectives est compatible avec différentes théories des émotions. La théorie que nous privilégions, présentée de manière détaillée et défendue ailleurs (Tappolet, 2000, 2016), est la théorie perceptuelle des émotions. Selon cette théorie, les émotions sont des expériences perceptuelles de propriétés évaluatives, c'est-à-dire des expériences perceptuelles qui représentent (de manière non conceptuelle)⁷ leurs objets intentionnels comme possédant des propriétés évaluatives spécifiques. Ainsi, un épisode de peur consiste en une expérience perceptuelle qui représente son objet comme étant effrayant, un épisode de joie consiste en une expérience perceptuelle qui représente son objet comme étant joyeux, un épisode de colère consiste en une expérience perceptuelle qui représente son objet comme étant offensant, et ainsi de suite.

Comme indiqué ci-dessus, nous pensons que les humeurs et les (dé)plaisirs sensoriels sont également des évaluations affectives. Comme les émotions, ce sont des expériences perceptuelles de propriétés évaluatives. Cependant, ils se distinguent des émotions par certains aspects importants. Prenons l'exemple d'une humeur telle que l'appréhension. Lorsque vous ressentez de l'appréhension, vous faites l'expérience d'un danger imminent. De manière un peu plus technique, vous faites l'expérience de l'instanciation de l'effrayant. En cela, l'appréhension n'est pas distincte de la peur. Nous pouvons exprimer cette idée en disant que l'appréhension et la peur sont toutes deux des expériences perceptuelles de la même propriété évaluative, à savoir l'effrayant. Cependant, contrairement à la peur, l'appréhension ne semble pas être dirigée vers quelque chose en particulier. Cela met en évidence une caractéristique plus générale des humeurs. Contrairement aux émotions, les humeurs ne semblent pas avoir d'objets intentionnels spécifiques. Cette idée a été élaborée de différentes manières dans la littérature. Par exemple, l'un d'entre nous a soutenu la thèse selon laquelle les humeurs ont des objets que l'individu est incapable d'identifier. Associé à une théorie perceptuelle, cela conduit à l'idée que les humeurs sont des expériences perceptuelles qui représentent des objets non identifiés comme possédant des propriétés évaluatives spécifiques (Rossi, 2021). D'autres points de vue sont néanmoins possibles. Par exemple, les humeurs peuvent être conçues comme ayant des objets intentionnels généralisés (Solomon, 1976/1993 ; Prinz 2004), pluriels (Siemer, 2009) ou modaux (Price, 2006 ; Tappolet, 2018). Pour l'instant, le seul point sur lequel nous nous engageons est que, si les humeurs sont une sorte d'expériences perceptuelles de propriétés évaluatives, comme les

émotions, elles diffèrent des émotions en ce qui a trait à leur objet intentionnel, qui est non spécifique.

Des considérations similaires s'appliquent aux plaisirs sensoriels (et aux déplaisirs sensoriels). Par plaisirs sensoriels, nous entendons toute expérience sensorielle qui possède la propriété d'être plaisante, c'est-à-dire toute expérience sensorielle plaisante. Ainsi, pour comprendre ce que sont les plaisirs sensoriels, il faut comprendre, premièrement, ce que sont les expériences sensorielles et, deuxièmement, ce que sont le plaisant et le déplaisant. Prenons par exemple le plaisir de déguster un café le matin. L'un d'entre nous a défendu la thèse évaluativiste suivante (Rossi, Manuscrit). Premièrement, une expérience sensorielle consiste en une expérience perceptuelle qui représente un objet (par exemple le café) comme possédant certaines qualités sensorielles (par exemple un goût particulier). Résumons cela en disant qu'une expérience sensorielle consiste en une expérience perceptuelle qui représente un « objet sensoriel ». Ensuite, une expérience sensorielle qui est *plaisante* consiste en ce qu'elle représente en plus l'objet sensoriel *comme agréable*⁸. Notons que, dans ce contexte, le caractère plaisant et le caractère agréable sont deux propriétés distinctes. Le caractère plaisant est une propriété phénoménologique de l'expérience sensorielle, tandis que le caractère agréable est une propriété évaluative dont l'objet sensoriel est représenté comme porteur (et qu'il peut véritablement posséder si la représentation est véridique – voir ci-dessous pour plus de détails). Plus précisément, le caractère agréable est une propriété évaluative *déterminable*, c'est-à-dire une propriété dotée de différentes spécifications déterminées. Cela signifie qu'il existe différentes façons pour un objet sensoriel d'être agréable. Par exemple, le goût du café peut être exquis, délicieux, réconfortant, et ainsi de suite. Une implication est que, alors que la classe des plaisirs sensoriels est unifiée par le fait que tous les plaisirs sensoriels ont l'agréable comme objet formel, différents types de plaisirs sensoriels peuvent être distingués par la manière particulière qu'ont leurs itérations d'être agréables. Cela dit, l'idée principale à retenir ici est que, dans la mesure où les plaisirs sensoriels consistent en des expériences perceptuelles qui représentent des objets sensoriels comme agréables (d'une manière déterminée), et dans la mesure où être agréable est une propriété évaluative, alors les plaisirs sensoriels comptent aussi comme des expériences perceptuelles de propriétés évaluatives. En cela, ils sont comme les émotions et les humeurs. Ils diffèrent cependant des émotions et des humeurs à deux égards. Le premier est qu'ils ont toujours un objet sensoriel comme objet intentionnel. Le second est que ce ne sont pas toutes les émotions et humeurs positives qui représentent des propriétés évaluatives consistant en des spécifications de l'agréable.

Avant de passer à notre prochaine thèse, quelques mots sur les valorisations s'avéreront utiles. En effet, comme nous l'avons vu, elles sont pertinentes pour déterminer la centralité dite « de source » des états constitutifs du bonheur. Historiquement, les théories les plus populaires ont conçu les valorisations soit comme une forme de désir de premier ou de second ordre (par exemple, valoriser x est désirer x ; ou valoriser x est désirer de désirer x), soit comme une forme de jugement de valeur (par exemple, valoriser x est juger que x a de la valeur). Cepen-

dant, ces deux approches font l'objet de sérieuses objections (voir Seidman, 2009 ; Scheffler, 2010 ; Svavarsdóttir, 2014, entre autres.) En conséquence, certains chercheurs se sont tournés vers une approche de la valorisation selon laquelle évaluer un élément x , c'est s'en soucier (par exemple, Seidman, 2009). Nous pensons qu'il s'agit d'un pas dans la bonne direction, mais que ce n'est pas encore entièrement satisfaisant. Le souci que l'on éprouve pour une personne ou un objet est un état mental qui appartient à la classe des sentiments. La classe des sentiments comprend toutefois d'autres états mentaux, tels que l'amour, la haine, la jalousie, etc. Ce qui unifie cette classe est le fait que tous ces états mentaux sont des dispositions émotionnelles à voies multiples, c'est-à-dire des dispositions à éprouver une variété d'émotions liées à l'objet intentionnel du sentiment. Ce qui importe pour nous ici, c'est que *tous* ces sentiments possèdent les caractéristiques qui sont typiquement associées aux valorisations. Ailleurs, nous avons ainsi proposé de considérer la notion de valorisation comme se référant à la classe des sentiments, plutôt qu'à un seul type d'état mental tel que le souci, et de considérer les sentiments comme différents types de valorisations (Rossi et Tappolet, Manuscrit-a).

En tant qu'états dispositionnels, les sentiments n'ont pas de dimension phénoménologique propre, au-delà de la phénoménologie des états affectifs occurrents qu'ils disposent un individu à ressentir. Ils ont, cependant, un contenu représentationnel semblable à celui des états affectifs occurrents. En particulier, ils ont un objet intentionnel, qui est représenté (de manière non conceptuelle) comme possédant une propriété évaluative particulière. Par exemple, lorsque vous aimez votre partenaire, votre amour a un objet intentionnel, c'est-à-dire votre partenaire, qui est représenté comme possédant une propriété évaluative particulière, c'est-à-dire la propriété d'être aimable. Leur caractère représentationnel distingue les sentiments d'autres types de dispositions affectives qui n'ont pas de contenu représentationnel. Pour marquer cette distinction, nous nous référons aux sentiments comme à des états affectifs *persistants*⁹. Un état persistant est tout état qui a un contenu représentationnel et qui, lorsqu'il est activé, génère des états occurrents. À titre d'exemple non affectif, considérez votre croyance que cet article s'intitule « Une théorie du bien-être comme bonheur approprié », que nous venons d'activer en écrivant cette ligne.

3. LE BONHEUR APPROPRIÉ

Faisons à présent le point. Nous avons affirmé que le bonheur consiste en une balance largement positive d'états affectifs tels que les émotions, les humeurs et les plaisirs sensoriels. Cette balance dépend, entre autres, de la relation entre ces états et les sentiments de l'individu. Deuxièmement, nous avons affirmé que les émotions, les humeurs et les plaisirs sensoriels sont des sortes d'expériences perceptuelles de propriétés évaluatives, et que les sentiments sont une sorte d'états évaluatifs persistants. Il découle de ces affirmations que le bonheur consiste en une balance largement positive d'expériences perceptuelles de propriétés évaluatives, qui est basée sur les évaluations persistantes de l'individu. Cette théorie a une implication immédiate pour notre compréhension du

bonheur. Le bonheur peut être conçu comme un état composite qui nous informe affectivement sur les valeurs que nous rencontrons dans notre vie. En ce sens, le bonheur compte comme une évaluation affective globale, qui tient compte de la perspective évaluative plus générale de l'individu.

Nous pouvons maintenant passer à notre quatrième thèse. Comme nous l'avons vu, les émotions, les humeurs et les plaisirs sensoriels sont des expériences perceptuelles qui représentent leurs objets intentionnels comme possédant des propriétés évaluatives spécifiques. En tant que types d'expériences perceptuelles, elles peuvent être évaluées comme étant appropriées ou non appropriées. Par *approprié*, nous entendons la même chose que « correct au niveau représentationnel ». Par conséquent, les émotions, les humeurs et les plaisirs sensoriels sont appropriés si et seulement si le monde est tel qu'ils le représentent. En d'autres termes, les émotions, les humeurs et les plaisirs sensoriels sont appropriés si et seulement si leurs objets possèdent réellement les propriétés évaluatives que ces états représentent comme étant les leurs. Considérons, à titre d'exemple, un épisode de peur ressenti à l'égard d'un chien. Comme nous l'avons vu, cet épisode consiste en une expérience perceptuelle qui représente le chien comme effrayant. Selon nous, un tel épisode de peur est approprié si et seulement s'il y a réellement un chien qui vous menace (plutôt que, par exemple, l'hologramme d'un chien) et si le chien est réellement effrayant (plutôt que, par exemple, un chiot inoffensif). Des considérations similaires s'appliquent aux humeurs et aux plaisirs sensoriels. Ajoutons que, selon notre théorie, les sentiments ont également des conditions d'adéquation. Considérons, par exemple, un sentiment tel que l'amour pour une personne particulière. On peut soutenir que ce sentiment peut également être évalué comme étant approprié ou non. Il est approprié à condition qu'une telle personne soit véritablement digne d'être aimée ou, de manière équivalente, « aimable ». Sinon, il est inadéquat, dans le sens où il n'est pas bien ajusté à la réalité.

Ensuite, notre théorie soutient que, dans la mesure où le bonheur est constitué par des états qui ont des conditions d'adéquation, le bonheur peut également être évalué comme étant approprié ou inapproprié. Pour commencer, nous pouvons dire que, pour que le bonheur soit approprié, il faut que les états qui le constituent soient appropriés, c'est-à-dire que leurs objets intentionnels possèdent effectivement les propriétés évaluatives qu'ils représentent comme étant les leurs. Puisque ces états sont typiquement basés sur des sentiments, qui doivent être pris en compte pour déterminer la mesure dans laquelle un individu est heureux, une autre condition doit être ajoutée pour que le bonheur soit approprié : il doit également être basé sur des sentiments appropriés. Cette théorie a deux implications. La première est que le bonheur peut être plus ou moins approprié, en fonction de la mesure dans laquelle ses constituants et leurs bases permanentes sont appropriés. La seconde implication est que, dans la mesure où le bonheur approprié consiste en une balance largement positive d'expériences perceptuelles *appropriées* de propriétés évaluatives, qui sont basées sur des évaluations persistantes *appropriées*, alors le bonheur approprié peut être caractérisé à la fois comme : a) une expérience largement positive d'éléments qui ont

une réelle valeur : b) une évaluation « globale » de la situation d'un individu (à un moment donné ou dans un intervalle de temps donné).

4. LE BIEN-ÊTRE COMME BONHEUR APPROPRIÉ

Cela nous amène à notre dernière thèse. Elle consiste simplement à dire que le bien-être n'est rien d'autre que le bonheur approprié. Nous pouvons résumer notre théorie de manière plus précise en nous référant à trois questions que toute théorie du bien-être doit aborder. Premièrement, *quels* éléments sont bons pour un individu, d'une manière qui ne soit pas instrumentale ? Deuxièmement, *qu'est-ce qui rend* ces éléments bons pour l'individu de manière non instrumentale ? Troisièmement, quelles dimensions sont pertinentes pour déterminer le *degré* de bien-être d'un individu ?

Selon notre théorie, le bien prudentiel de base est le bonheur approprié. C'est en cela que consiste le bien-être au niveau le plus fondamental. Le bonheur approprié constitue le bien-être en vertu du fait qu'il s'agit d'une expérience affective du bien (*simpliciter*). C'est ce qui le rend non instrumentalement bon *pour* un individu. Quant aux variables qui déterminent le degré de bien-être d'un individu, il s'agit de l'intensité phénoménale des états constitutifs du bonheur approprié et de leur centralité de source. À la lumière de ce que nous avons dit précédemment, il s'ensuit que la mesure dans laquelle la vie d'un individu se déroule bien pour l'individu qui la vit est déterminée par la mesure dans laquelle cet individu fait l'expérience affective d'éléments qui ont réellement de la valeur (*simpliciter*), et par la mesure dans laquelle ces expériences sont basées sur les évaluations affectives persistantes de l'individu et y correspondent.

Il convient de noter une complication. Dans certains cas, les états constitutifs du bonheur d'un individu peuvent être appropriés au niveau de leurs objets, c'est-à-dire dans la mesure où ils représentent comme ayant de la valeur des objets qui en ont réellement, mais pas sur le plan de leur intensité, c'est-à-dire dans la mesure où ils représentent comme ayant un certain degré de valeur des objets qui ont un degré de valeur différent. De même, dans d'autres cas, les états constitutifs du bonheur d'un individu peuvent être appropriés au niveau de leurs objets et de leur intensité, mais ils peuvent être basés sur des sentiments qui ne sont pas appropriés ou seulement partiellement appropriés. Par exemple, cela peut se produire lorsqu'un individu éprouve correctement de la joie à l'occasion d'un événement *x*, mais que sa joie provient de la valorisation d'éléments d'une classe donnée, alors que tous les membres de cette classe ne méritent pas réellement de la joie. Dans tous ces cas, les états constitutifs du bonheur de l'individu ne sont que partiellement appropriés. Cela introduit certaines difficultés intrigantes dans l'estimation du degré global de bonheur adéquat, que nous n'avons pas, malheureusement, la place pour explorer dans cet article.

Il ressort de ce qui précède que ce que nous proposons est une théorie hybride du bien-être basée sur le bonheur. Selon notre approche, le bien-être consiste en une sorte de bonheur adéquatement qualifié, à savoir le bonheur *approprié*. Dans

l'introduction, nous avons mentionné que les théories hybrides fondées sur le bonheur sont confrontées à un défi explicatif. Elles doivent fournir une explication convaincante, et non *ad hoc*, de la raison pour laquelle le bonheur doit être qualifié d'une manière particulière pour que la vie d'un individu se déroule bien pour lui. Considérons, par exemple, les théories selon lesquelles le bonheur doit être mérité ou authentique pour bénéficier à un individu. On peut affirmer que l'une ou l'autre de ces qualifications est nécessaire pour rendre la théorie du bien-être qui en résulte convenable sur le plan descriptif et/ou normatif. En d'autres termes, ces qualifications sont nécessaires pour atteindre l'équilibre réfléchi. Pourtant, le défi explicatif demeure. Qualifier le bonheur comme le font ces théories permet de *rendre compte* des données et de respecter nos jugements concernant des cas particuliers. Mais peut-il également *expliquer* les données ?

Le même défi s'applique à notre théorie. En effet, d'un point de vue extensionnel, notre théorie appartient à la famille des théories qui considèrent que le bien-être consiste en une expérience du bien (*simpliciter*). Ces théories sont souvent critiquées comme étant une combinaison *ad hoc* des théories de l'état mental et des théories objectivistes du bien-être. Elles reprennent à leur compte la thèse selon laquelle le bien-être nécessite des expériences affectives, une thèse qui se rapporte à la première théorie, et la thèse selon laquelle le bien-être nécessite la présence d'objets ayant réellement de la valeur objective, qui se rapporte à la seconde. Toutefois, selon l'objection, cette fusion n'est pas justifiée de manière indépendante.

Nous souhaitons conclure cette section en esquisant une réponse à cette objection. Ce faisant, nous voulons montrer que notre théorie présente un avantage important par rapport aux autres théories hybrides. Il nous semble que, par rapport à ces dernières théories, notre théorie apporte une couche supplémentaire à l'explication. Rappelons que, selon notre théorie, le bonheur consiste en une balance largement positive d'expériences perceptuelles de propriétés évaluatives. En tant que tel, le bonheur fournit à l'individu à la fois une expérience d'éléments de valeur et une évaluation de sa situation. Cependant, comme nous l'avons mentionné plus haut, le bonheur peut accomplir ces tâches correctement ou incorrectement. Cette idée constitue la base de notre réponse à l'objection précédente. Selon notre théorie, le bonheur détermine le bien-être en vertu du fait qu'il est une expérience affective du bien, où cette expérience fournit également à l'individu une évaluation globale de sa situation. Dans la mesure où nous lions le bien-être à ces caractéristiques, il est naturel de penser que le bien-être dépend finalement du fait que le bonheur accomplisse ces tâches *correctement*. Ainsi, le qualificatif « approprié » attaché au bonheur n'est pas un ajout motivé par la seule nécessité de se conformer aux données. Il ne s'agit pas non plus d'un élément emprunté à une autre théorie, qui ne concerne pas le bonheur. Au contraire, l'ajout de cette qualification est un moyen de fournir une explication *complète* de ce qui fait du bonheur un élément constitutif du bien-être.

5. CONCLUSION

À première vue, il pourrait sembler que, pour défendre pleinement notre théorie du bien-être, nous devons la comparer à toutes les principales théories concurrentes. Cette tâche semble décourageante. Mais il existe une autre stratégie, qui pourrait rendre notre travail plus aisé. Elle consiste à prendre au sérieux le lien intuitif qui existe entre le bien-être et le bonheur. On peut soutenir qu'il s'agit d'une raison *prima facie* de prendre les théories du bien-être fondées sur le bonheur comme point de départ par défaut de l'analyse du bien-être. Si cela est vrai, il suffit de montrer que notre théorie est immunisée contre les objections soulevées à l'encontre des autres théories du bien-être fondées sur le bonheur, pour qu'elle soit considérée comme une théorie plausible du bien-être. Combiné aux avantages mentionnés dans la section précédente, cela montrerait en effet que notre théorie est une théorie prometteuse du bien-être.

En guise de conclusion, nous souhaitons fournir une liste des objections qui ont été soulevées contre les théories traditionnelles basées sur le bonheur, et auxquelles notre théorie doit répondre. Nous pouvons énumérer cinq objections principales. La première est « l'objection de la superficialité ». Selon cette objection, le bonheur est un état psychologique trop superficiel ou trop éphémère pour être le constituant ultime, ou même simplement un constituant, du bien-être d'un individu. La seconde est « l'objection de l'absence de valeur ». Selon cette objection, le bonheur ne peut être un élément constitutif du bien-être, car on peut tirer du bonheur d'objets dépourvus de toute valeur réelle, alors qu'on ne dirait pas que tirer du bonheur d'une activité vaine, comme celle de compter les brins d'herbe sur une pelouse, contribue au bien-être. La troisième est « l'objection de l'inauthenticité ». Elle affirme que le bien-être ne peut être fondé sur le bonheur, car celui-ci peut être inauthentique. Il peut, par exemple, résulter d'un lavage de cerveau. Lorsque le bonheur est inauthentique, il ne semble pas contribuer au bien-être d'un individu. La quatrième objection, que l'on peut appeler « l'objection de la passivité », affirme que les théories fondées sur le bonheur sont trop « passives », car elles admettent la possibilité qu'un individu puisse tirer du bien-être de son bonheur sans exercer sa capacité d'agentivité. Ce résultat est considéré comme contre-intuitif. La dernière objection, qui peut être nommée « l'objection du mal-être », affirme que, selon notre théorie, si le malheur est totalement inapproprié, alors la vie d'un individu ne se passe pas du tout mal pour lui. Il semble certain, cependant, que cela ne peut être vrai, car le malheur a une phénoménologie négative et cela semble rendre le malheur mauvais pour l'individu qui l'éprouve, même si son malheur est par ailleurs inapproprié.

Nous pensons que notre théorie a les moyens d'éviter toutes ces objections – et nous avons d'ailleurs proposé divers arguments en ce sens ailleurs¹⁰. Nous espérons que cet article a offert au moins une première idée des atouts de notre théorie.

REMERCIEMENTS :

Nous tenons à remercier les évaluateurs pour leurs objections et leurs remarques. De plus, nous aimerons remercier les participants à la conférence SIFA 2018, à l'Université du Piémont oriental, les participants à l'atelier « Bien-être et états affectifs », à l'Université Clermont Auvergne, les participants à l'atelier « Perspectives sur le bien-être », au Centre de recherche en éthique (Montréal), et les participants à la « Fittingness Conference », pour leurs commentaires utiles. Nous sommes particulièrement reconnaissants à Selim Berker pour ses généreux commentaires sur une version antérieure de ce document. Nos travaux ont bénéficié du soutien du CRSH, que nous tenons à remercier.

NOTES

- ¹ Cet article est basé sur un chapitre intitulé « Well-Being as Fitting Happiness », à paraître dans un volume sur la notion d'adéquation (*fittingness*) pour Oxford University Press.
- ² Par conséquent, lorsque nous disons qu'une personne a un bien-être élevé, nous voulons dire que la vie de cette personne se déroule bien pour elle d'un point de vue prudentiel. Au contraire, lorsque nous disons qu'une personne est heureuse, nous voulons dire qu'elle est dans un état mental positif, ou une combinaison d'états mentaux positifs.
- ³ En anglais, nous utilisons le terme *fitting*.
- ⁴ Voir Rossi et Tappolet (Manuscrit-b), (Manuscrit-c), (Manuscrit-d).
- ⁵ Pour cette caractérisation des valorisations, voir, entre autres, Seidman, 2009 ; Scheffler, 2010 ; Svavarsdóttir, 2014.
- ⁶ Voir (Teroni, 2007)
- ⁷ Au sujet de la notion de contenu non conceptuel, voir Tappolet, 2020. Par souci de simplicité, nous omettrons cette qualification dans ce qui suit.
- ⁸ En anglais, nous utilisons le terme *pleasurable*, que nous contrastons avec *pleasant*.
- ⁹ En anglais, nous parlons de *standing affective states*.
- ¹⁰ Voir Rossi et Tappolet (Manuscrit-b), (Manuscrit-c), (Manuscrit-d).

BIBLIOGRAPHIE

Haybron, Daniel M., *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. New York, Oxford University Press, 2008.

Price, Carolyn, « Affect Without Object: Moods and Objectless Emotions », *European Journal of Analytic Philosophy*, vol. 2, no. 1, 2006, p. 49-68.

Prinz, Jesse J., *Gut Reactions: A Perceptual Theory of the Emotions*, New York, Oxford University Press, 2004.

Rossi, Mauro, « Happiness, Pleasures, and Emotions », *Philosophical Psychology*, vol. 31, no. 6, 2018, p. 898-919.

———, « A Perceptual Theory of Moods », *Synthese*, vol. 198, no. 8, 2021, p. 7119-7147.

———, « An Evaluativist Theory of Pleasure », Manuscrit.

Rossi, Mauro et Christine Tappolet, *An Affective Theory of Happiness*, Manuscrit-a.

———, « Happiness as an Affective Evaluation », Manuscrit-b.

———, « Ill-being and Fitting Unhappiness », Manuscrit-c.

———, « Well-being as Fitting Happiness », Manuscrit-d.

Scheffler, Samuel, *Equality and Tradition: Questions of Value in Moral and Political Theory*, New York, Oxford University Press, 2010.

Seidman, Jeffrey, « Valuing and Caring », *Theoria*, vol. 75, no. 4, 2009, p. 272-303.

Siemer, M., « Mood Experience: Implications of a Dispositional Theory of Moods », *Emotion Review*, vol. 1, no. 3, 2009, p. 256-263.

Solomon, Robert C., *The Passions: Emotions and the Meaning of Life*, Indianapolis, Indiana, Hackett Publishing Company, 1976/1993..

Svavarsdóttir, Sigrún, « Having Value and Being Worth Valuing », *Journal of Philosophy*, vol. 111, no. 2, 2014, p. 84-109.

Tappolet, Christine, *Émotions et valeurs*, Paris, Presses Universitaires de France, 2000.

———, *Emotions, Values, and Agency*, Oxford, Oxford University Press, 2016.

———, « The Metaphysics of Moods », in Hichem Naar et Fabrice Teroni (dir.), *The Ontology of Emotions*, Cambridge, Cambridge University Press, 2018, p. 169-186.

———, « Emotions Inside Out: The Nonconceptual Content of Emotions », in Christoph Demmerling et Dirk Schroeder (dir.), *Concepts in Thought, Action, and Emotion: New Essays*, New York, Routledge, 2020, p. 257-276.

Teroni, Fabrice, « Emotions and Formal Objects », *Dialectica*, vol. 61, no. 3, 2007, p. 395-415.

VIRTUE, HAPPINESS, AND EMOTION

ANTTI KAUPPINEN

PRACTICAL PHILOSOPHY, UNIVERSITY OF HELSINKI

ABSTRACT:

Philosophers have tried very hard to show that we must be virtuous to be happy. But as long as we stick to the modern understanding of happiness as something experienced by a subject – and I argue against contemporary eudaimonists that we should indeed do so – there can at best exist a contingent causal connection between virtue and happiness. Nevertheless, we have good reason to think that being virtuous is non-accidentally conducive to happiness. Why? First, happiness is roughly the experiential condition of enjoying predominantly positive affective phenomenal states concerning things that are subjectively important to us. I argue that this straightforward sentimentalism about happiness has several advantages over Daniel Haybron's emotional condition account. Second, insofar as we're virtuous, we can correctly identify what is worth doing in our particular situation and will skillfully pursue it. At the same time, we're not bothered by things that are not worth caring or worrying about. Consequently, virtuous people are likely to enjoy central positive emotions related to success and approval by others, and avoid common negative emotions related to social comparison or avarice. While their happiness is still in part a matter of luck, it is such to a lesser degree than for the rest of us.

RÉSUMÉ :

Les philosophes se sont efforcés de montrer que nous devons être vertueux pour être heureux. Mais tant que nous nous en tenons à la compréhension moderne du bonheur comme quelque chose de vécu par un sujet – et je soutiens contre les eudaimonistes contemporains que nous devrions effectivement le faire – il peut au mieux exister un lien de causalité contingent entre la vertu et le bonheur. Néanmoins, nous avons de bonnes raisons de penser qu'être vertueux est non accidentellement propice au bonheur. Pourquoi? Premièrement, le bonheur est, sommairement, la condition expérientielle de jouir d'états phénoménaux affectifs principalement positifs concernant des choses qui sont subjectivement importantes pour nous. Je soutiens que ce sentimentalisme à propos du bonheur a plusieurs avantages par rapport à la conception de Daniel Haybron comme état émotionnel. Deuxièmement, dans la mesure où nous sommes vertueux, nous pouvons identifier correctement ce qui vaut la peine d'être fait dans notre situation particulière et nous la poursuivrons habilement. En même temps, nous ne sommes pas dérangés par des choses qui ne valent pas la peine d'être prises en charge ou de s'inquiéter. Par conséquent, les personnes vertueuses sont susceptibles de profiter d'émotions positives centrales liées au succès et à l'approbation des autres, et d'éviter les émotions négatives communes liées à la comparaison sociale ou à l'avarice. Bien que leur bonheur soit encore en partie une question de chance, il l'est dans une moindre mesure que pour le reste d'entre nous.

INTRODUCTION

If movies are to be believed, villains, by and large, do not thrive—morally bad people lead lives that are in some way marred and unhappy, precisely because they're morally bad. The idea that virtue is necessary if not sufficient for happiness is, of course, one of the most persistent themes in European ethical thought from Socrates to Hume, at least. One surprising thing about this persistence is that it survived major changes in conceptions of both virtue and happiness itself. With the revival of ancient virtue ethics in the latter half of the twentieth century, the thought that happiness requires virtue has found at least tentative defenders even in our own day, such as Philippa Foot (2000), Julia Annas (2011), and Daniel Russell (2012). According to at least one study, it also resonates with many ordinary people's conception of happiness (Philips, Nyholm, and Liao, 2014).

Here's one way we can put the idea while taking into account the fact that both virtue and happiness are a matter of degree:

Virtue Necessity

Possessing the full set of ethical virtues of character to a high degree is necessary for a high level of happiness.

Given the way we tend to understand happiness these days, it's easy to dismiss Virtue Necessity as some kind of wishful thinking—wouldn't it be great if being good would be in everyone's self-interest? Nevertheless, I propose to take it seriously enough to examine the case that contemporary Aristotelians make for it. I will argue, though, that they end up stretching the notion of happiness so far that it's no longer recognizable as such. In consequence, there can be at best a *causal* connection between virtue and happiness. But perhaps this causal connection isn't quite a matter of chance, given the best way to understand of each. Maybe the following kind of interesting generalization still holds:

Virtue Conduciveness

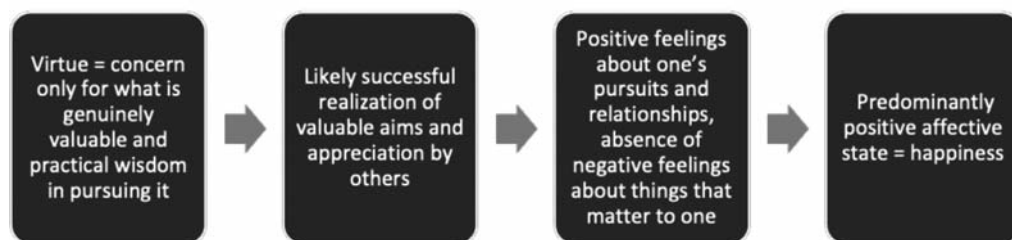
Other things being equal, possessing a full set of ethical virtues of character to a high degree and manifesting them in action increases happiness relative to acting from less complete or absent virtue.

In this paper, I'm going to argue that there are good *a priori* reasons to think that Virtue Conduciveness is true, at least for most people, and also that we can specify the *ceteris paribus* clause enough to say that it's not *trivially* true.

I'll make the case as follows. In the first section, I'll introduce the notion of virtue as I understand it. The conception I'll work with will be broadly Aristotelian: a virtue is a multitrack disposition to recognize reasons pertaining to a particular good and to respond to those reasons correctly in terms of emotion, deliberation, and action. It is crucial for my argument that virtue involve both (a) emotional concern (only) for what is genuinely valuable and (b) skill at ration-

ally pursuing related aims. Further, different virtues form a mutually supportive system. Second, as I said, I'll reject recent neoclassical defences of eudaimonistic conceptions of happiness. Their basic move is to start from a formal conception of our final end, which they identify with happiness, and argue that what best fulfills this happiness role is (or can be) a life of virtuous activity. I argue that no conception of happiness that meets reasonable minimal constraints is suitable to play the role of a unique final end in practical reasoning, which we could instead identify with something like "the good life" (Haybron) or the most choiceworthy life.

Third, I'll introduce my version of a sentimentalist account of happiness, according to which we are happy, roughly, to the extent that the phenomenal character of our experiences of the things that are subjectively important to us is predominantly positive in valence. I argue that this view of happiness as positive significance-adjusted affective balance meets more of the reasonable criteria for an account of the nature of happiness than Dan Haybron's (2008) industry-standard view. Assuming sentimentalism about happiness, I'll argue that, given the nature of virtue and common human concerns, we have good reason to think that Virtue Conduciveness is true. Here's a rough picture of the relevant mechanism:



The idea is that insofar as we're virtuous we'll skillfully pursue realistic, genuinely valuable aims—not only prospective ones of realizing future-directed goals, but also reflexive ones of performing worthwhile activities well—that take into account the interests and rights of people around us. At the same time, we lack desires for what doesn't really matter, like having a newer car than our neighbour. Given these facts about our cares and practical reasoning, it's likely that we're successful and are in good standing with people around us, which in turns fosters positive affect, while at the same time many known sources of negative affect are absent. So, virtuous outlook and activity are apt to *indirectly* bring about happiness in affective terms. In the final section, I'll compare and contrast this view with Mauro Rossi and Christine Tappolet's (2016; forthcoming) recent sentimentalist defence of noncontingent links among virtue, happiness, and well-being that bears some important similarities to the view defended here.

1. VIRTUE, EMOTION, AND REASON

To begin with, let's briefly consider the nature of virtue and, more specifically, ethical virtues of character, which contrast with intellectual virtues like theoret-

ical reasoning Let's start with a tentative list of some central putative virtues: justice, honesty, trustworthiness, generosity, kindness, temperance, tenacity, and courage. These are all morally admirable traits when manifest to a great degree. (It may well be that traits like wit or innovativeness are virtues in a broader sense, as philosophers like Hume like to think—they are admirable in some nonmoral way.) But what is it to have such a trait? Suppose that Bernie is honest. Everyone agrees that it follows that he tends to tell the truth even when it's inconvenient to him and that he can be trusted not to take what's not his. But such behaviour is neither sufficient nor necessary for honesty, since people who lack the virtue might do the same things, and, in appropriate circumstances, even honest people may arguably lie. What is distinctive about honest people is rather to be found in their recognition of reasons to be truthful and their psychological response to them.

According to the standard Aristotelian view, a full virtue is a complex disposition to have integrated cognitive, conative, and affective responses that are situationally appropriate. Bernie's honesty involves, first, the ability or skill to tell when telling the truth is the right thing to do. This skill requires a grasp of the point of telling the truth and, according to many, the ability to articulate reasons why something is the right thing to do (Annas, 2011, p. 19). As people like Alison Hills (2016) point out, it involves *understanding why* one should tell the truth rather than just *knowing* what one should do. Having to account for our choices by giving reasons forces us to reflect and go beyond what we learn from the behaviour of others around us with respect to, say, truth telling (Annas, 2011, p. 54).

In addition to this cognitive aspect, virtue has a *conative* part: Bernie has a desire to tell the truth when he considers it appropriate. Importantly, virtue also consists in *affective* dispositions, since it is a state or disposition (*hexis*) to feel pleasure, pain, anger, pity, and other emotions "at the right time, about the right things, towards the right people, for the right end, and in the right way" (NE 1106b: 30). So Bernie takes some pleasure in doing the right thing and would be pained and ashamed if he happened to mislead someone. Similarly, a generous person not only gives the right amount—neither excessively nor deficiently—but also does so without displeasure or resentment. Such feelings would manifest disproportionate concern for possessing the good thing in question instead of giving it to someone who needs it more. This makes for a difference between the virtuous and the merely continent or self-controlled, who do the right thing as a result of overcoming internal opposition. Fully honest people aren't tempted to lie, so they don't need to exercise self-control in order to be truthful (NE 1152a: 135). (To be sure, the picture is not so straightforward with virtues like temperance and courage, which are precisely virtues *of* self-control.) If an honest person is internally conflicted, for example, it's because there are genuinely conflicting reasons for and against telling the truth, so there may be a moral residue even if one does what's all-things-considered best.

The idea of the virtuous not being tempted by the bad ties in with one of the emphases of the Stoic tradition: the sage is distinguished by not caring about worldly things. The Stoics tended to take this to an extreme, holding that only virtue is good and vice bad, while everything else is strictly speaking indifferent (even if things like nourishment are “preferred indifferents”). I’m not convinced by the reasons they have offered for this extreme revisionism about value (see, e.g., Brennan, 2005, whose interpretation emphasizes that only virtuous activity actualizes human nature, which in turn determines our good). Nor am I even convinced of the coherence of the view (what makes justice a virtue if it doesn’t matter who has what?). But it *is* a real insight that an important part of being virtuous is not being too concerned with what doesn’t really matter.

According to the Aristotelian picture, affective dispositions are thus an important aspect of virtue. This is not in the first instance because of their motivational role: according to many, a correct perception of a reason to act is by itself sufficient to motivate the virtuous person to act. John McDowell, for example, argued that one can’t really conceive of a situation in virtue terms without also being motivated to act: “The relevant conceptions are not so much as possessed except by those whose wills are influenced appropriately” (McDowell, 1978, p. 87). Arguably, this point generalizes to appropriate emotional responses: *ceteris paribus*, you don’t really conceive of someone’s action as being despicable if you’re not at least disposed to despise or otherwise negatively react to it. Our emotions constitute, in part, our subjective perspective on the world; if we lack appropriate emotional responses when something of real value is at stake, we’re not sufficiently responsive to the value. As Thomas Hurka (2000) says, it is intrinsically good to love what is intrinsically good for itself and to hate what is intrinsically bad. While a part of what it is to value truth, for example, is to perceive a strong reason against lying for your own advantage, there is more to valuing truth than perceiving reasons. Your valuing truth is also manifest in your positive emotional response to truth telling for its own sake and in your disposition to be ashamed about lying. Someone who lacks such dispositions has a split or incomplete evaluative perspective, even if that person does the right thing. So, against Martha Nussbaum’s (2016) criticism of anger, I’ve argued that if we value justice, we must be angry about injustice when we stand in a suitable relation to it (Kauppinen, 2018). And since we do value justice if we’re virtuous, we are sometimes angry if we’re virtuous. For similar reasons, the courageous person who values the security of innocents takes pleasure in actions that promote security, even if these involve risk to herself or himself. Such person’s affective and cognitive perspectives are unified.

Having the right sort of emotions and desires means that one has the right sort of ends, but that doesn’t suffice for virtue yet—you still need to figure out what to actually do in your particular situation to promote the good. For Aristotle himself, virtues of character can be had without practical wisdom (*phronesis*), while practical wisdom can’t be had without virtues of character (NE 1144b: 118). Practical wisdom is thus something like skillful practical reasoning in

pursuit of situationally appropriate aims, distinct from mere cleverness or intelligence in pursuit of any goal whatever. Consider generosity, for example. As debates around effective altruism have shown, it's not a trivial question how to best give to others—should I, for example, work within the capitalist system to gain wealth that I can share with others, or work to change the system? It evidently takes practical reasoning to get at what to give (money, time, advice, attention, etc.), to whom, when, and how.

The debate about effective altruism also illustrates how virtues interact with each other: the values to which justice and generosity are responsive may be in tension with each other. Defenders of reciprocity of virtue from Aristotle to McDowell (1979) hold that virtues must come as a package, since in practical reasoning we need to take into account all sorts of ethically relevant considerations—it takes more than generosity to give in the right way, or, alternatively put, you can't have the full virtue of generosity without other virtues. We do, of course, sometimes say that someone is honest without being generous, or that loyalty can conflict with justice. It's even more natural to think that people can have executive virtues like courage and tenacity that bear on how they pursue whatever ends they may have. Defenders of reciprocity will thus maintain that the kind of honesty you can have without generosity is defective, even if most of the time it leads you to do the right thing, and that the kind of loyalty or courage you can have without justice is a simulacrum of virtue. I formulated Virtue Necessity and Virtue Conduciveness above in terms of a "full set of virtues," but, given the mutual dependence among virtues, it's not unreasonable to talk about "virtue" in the singular either.

Supposing that something like the above conception is correct, which character traits are virtues? As Philippa Foot (1978) emphasized, virtues of character can be seen as *corrective* of common human moral and prudential defects. That's why they're praiseworthy. For example, we tend to give too much weight to our own safety in comparison to other goods, and courage corrects for this defect. To say this is of course not to explain how much weight we should give to our own safety and why. That's a matter for first-order ethical theory or perhaps more broadly a theory of the good life, all things considered (Haybron, Manuscript). I want to remain neutral on these substantive views, though my own take is that theories of value and right action are explanatorily prior to any account of what makes a character trait a virtue (cf. Hurka, 2000)—embracing the importance of the notion of virtue doesn't entail accepting any form of virtue ethics, which would reverse the explanatory priority. But focusing on human defects is useful here, since it's easy enough for people holding different first-order views to agree on our characteristic shortcomings: we tend to be too selfish and short-sighted, and to be biased in favour of our own kind, consumed by desire for approval, and tempted by pleasure or deterred by apparent danger. Modern (social) psychology has added considerably to the list of human defects: we're influenced by framing, priming, and the status quo; seduced by choice architecture, and subtly biased against outgroups and those who are statistically abnormal. We are also sensitive to social comparison and surprisingly conform-

ist, to mention just a few things (see, e.g., Thaler and Sunstein (2008) and Kahneman (2011) for influential overviews of these studies). I don't think this means that there is no such thing as Aristotelian virtue (*pace* Doris, 2002). It's just that it's even harder to be fully virtuous than we may have thought.

For the purpose of discussing Virtue Necessity, I'll simply take for granted the common-sense first-order view that the right thing to do sometimes involves putting ourselves in harm's way, letting go of something we really want, risking our health, and even giving priority to the needs of a complete stranger over those of a friend or ourselves. This gives rise to a simple and familiar challenge: how can we be sure that virtue will make us happy (or happier) when it appears to common sense that it may on occasion lead to utter misery?

2. WHY HAPPINESS ISN'T VIRTUOUS ACTIVITY

Given that Virtue Necessity is counterintuitive, there's a temptation to qualify it by saying that virtue is necessary for *true* happiness, where "true" is meant to convey that the ordinary understanding is mistaken or at least ambiguous. Here is how the argument in these terms might go:

The Eudaimonist Argument for Virtue Necessity

1. True happiness consists in virtuous activity with which one is subjectively satisfied.
2. Virtue is necessary for engaging in virtuous activity with which one is subjectively satisfied.
3. So, virtue is necessary for true happiness.

The first premise evidently goes against the way most of us think of happiness these days, though eudaimonists like to point to some contemporary usages of "happiness" that require meeting some objective standards in addition to feeling subjectively satisfied (Kraut, 1979). Either way, eudaimonists are right to insist that this isn't decisive. It's certainly possible that most people are indeed mistaken about the nature of happiness. At the same time, there's a risk of changing the topic if supposedly "true" happiness is insufficiently linked to our actual shared concept of happiness. To ensure this doesn't happen, we need to identify some key conceptual connections between happiness and other things—in Michael Smith's (1994) terms, some platitudes about happiness that had better come out true. This will give us a job description for happiness against which we can measure competing accounts without relying only on intuitions about who's happy and who isn't.

The job description that eudaimonists focus on highlights the putative *deliberative* role of happiness as the final end. Here is a sketch of the argument:

The Final End Argument

1. Deliberation requires a unique final end that is complete and self-sufficient.

2. Of the ends that serve as starting points for deliberation, only happiness is complete and self-sufficient, provided that its initially indeterminate content is suitably specified—“happiness” is a placeholder for whatever the final end consists in.
3. In order to serve as a complete and self-sufficient end, happiness must consist in virtuous activity with which one is subjectively satisfied.
4. So, happiness consists in virtuous activity.

While this type of argument derives from Aristotle, I’m going to focus on contemporary formulations by Julia Annas and Daniel Russell. Let’s start by examining the first premise about the nature of deliberation. Julia Annas (1993; 2011) emphasizes that for ancient ethical theories, the key starting point is reflection on our lives as a whole: we stand back from our ongoing concerns and ask ourselves what the best way for us to go on living would be. As rational agents, we do things *for the sake of* some end, which we regard as good. These ends form a hierarchy, in which realizing one end is either a means to or a way of realizing another. Perhaps we exercise for the sake of health or study for the sake of getting a job or go to the movies for the sake of having fun. Here we take maintaining health, getting a job, and having fun to be higher-level ends for the sake of which we do many other things. But why be healthy, have a job, and have fun? And what to do if these ends conflict? According to Aristotelians, to answer such questions rationally, we need a further end for the sake of which we seek both to be healthy and have fun, say, and we need to consider whether it is better promoted or realized by the health-promoting or the fun-constituting option. If there is more than one end at this level of the hierarchy, we keep going in the same way until we reach the final end or good for the sake of which other goods are pursued, and which we don’t pursue for the sake of a yet further end. This will then be our highest good (*NE* 1094a).

As Aristotle says, this final end has certain formal characteristics: it’s *complete*, in the sense of being something pursued for its own sake and not also for the sake of something else, and *self-sufficient*, in the sense of needing nothing to be added to it for a good life, at least provided that certain background conditions, like sufficient health and wealth, are in place (*NE* 1097a-b: 10–11). Being satisfied with our lives is part of what’s required for self-sufficiency, since otherwise there would be something that could be added to the final good to make it better.

The next step is the claim that *happiness* is the final good for the sake of which other goods are pursued and which is itself never pursued for anything else. Aristotle himself says that everyone agrees that happiness (or *eudaimonia*) fits the bill (*NE* 1095a: 5), and Annas says that “once we see the need for a final end, an end we pursue in some way in every action, there is no better specification than happiness that we can come up with for it” (1993, p. 332). In a way, this doesn’t help much, since the nature of happiness is itself in dispute. Indeed, Annas maintains that “happiness” is in fact largely a placeholder for whatever has the formal characteristics of a final end: “Happiness in ancient theories is

given its sense by the role it plays; and the most important role it plays is that of an obvious, but thin, specification of the final good” (1993, p. 46). So the hypothesis, if you will, is that happiness is *whatever* best answers the job description of a final good for beings like us who are asking the question. Methodologically, we might say that this is the conceptual role that we hold fixed in figuring out what happiness consists in.

Third, for happiness to be the organizing and self-sufficient goal for us, it can’t be a state like pleasure, because then it couldn’t serve as the focus of large-scale deliberation about our lives as a whole. Here Aristotelians draw on the *endoxa*, the reputable opinions of all, most, or the wise, to argue that something like pleasure can’t be the focus of the best kind of life for a human being, because pursuing it as the highest good would result in “a life fit only for cattle” (*NE* 1095b: 7). Given *our* nature, our final end must consist in *rational activity* of some sort. Eudaimonists take such activity to be *virtuous* activity, which involves the use of practical wisdom as well as virtues of character, possibly because virtue is just the name for the kind of excellence that helps something perform its function or distinctive job (*ergon*), and our job is to engage in rational activity (*NE* 1097b). The conclusion then follows: being our final end, happiness consists, at least in part, in virtuous activity. This then feeds into the main eudaimonist argument for Virtue Necessity.

What should we make of this picture? I think there are several hurdles in the way of identifying happiness as whatever plays the deliberative role of final end, even if we allow that our ordinary notion of happiness is “weak and flexible,” as Annas puts it (1993, p. 129), and we are willing to revise it. At the very least, our happiness is still something that is good *for us*. More precisely, my happiness is nonderivatively good for me, so anything that is constitutive of it must be nonderivatively good for me. But is it really the case that everything we have reason to pursue for its own sake and not for the sake of anything else, and for the sake of which we pursue other things, must be good *for us*? Doesn’t virtue, in particular, require pursuing things like justice for the sake of *others*, and perhaps some *impersonal values*, not for the sake of ourselves?

This is, of course, an ancient worry. Russell puts it very nicely: “In the one direction, we look for a good that could make sense of all the ends we have reason to pursue; but surely, the worry goes, many of these ends cannot plausibly be understood as being for the sake of one’s own good life” (2012, p. 24). Unsurprisingly, then, eudaimonists have standard responses. One basic idea is to distinguish between developmental stages or levels of explanation. We might *initially* cultivate virtue or, for that matter, friendship for the sake of leading the best kind of life for ourselves, but *subsequently* pursue the good of others for its own sake. By forming personal relationships, we make the good of others part of our own good. Russell says that “eudaimonism is a view about how we adopt our ends—for the sake of a good life—but it does not say that the ends we adopt should be self-serving” (2012, p. 26).

Even if we accept this developmental idea, a dilemma still remains. After deliberation and developing virtue and relationships, *either* we have reason to pursue the good of others its own sake and not for the sake of our own good *or* we continue to have reason to pursue the good of others for the sake of our own good. As one might put it, the problem isn't just that we're (perhaps initially) *motivated* to pursue all other ends for the sake of our own happiness, but also that our *reasons* for pursuing all other ends derive from their relation to our own happiness. The dilemma is then that if our reasons for acting for the sake of others don't derive from our own happiness, happiness isn't the final end of practical reasoning, but if they do, the justificatory structure of practical reasoning remains fundamentally egoistic, despite all the protestations of eudaimonists (such as Annas, 1993, p. 322).

Now, eudaimonists might argue that the first horn of the dilemma is based on an illusory opposition between the good of others and our own good. For the virtuous and the caring, they might say, the two are one and the same thing (cf. Foot, 2001, p. 102). I don't deny that we can make the good of *some* others part of our own good. That is surely what happens in some personal relationships. It's better for *me* to lose a limb than to lose a child, so that giving up a leg for a child would not be a genuine sacrifice any more than going to the dentist for a painful operation now for the sake of avoiding an even greater pain in the future is one. But not all virtues are exercised in that context, and, even there, genuine sacrifices are possible.

This leaves the eudaimonist stranded on the second horn: holding that virtuous people act ultimately for the sake of their own happiness after all, at least in the sense that their reasons for doing so derive from promoting or realizing their happiness. Somehow, our own good (which includes subjective satisfaction) continues to shape our final end. To make this palatable, the eudaimonist may insist that this doesn't presuppose that virtuous people somehow have their own self-interest in mind while deliberating or acting. Being "for the sake of" their own eudaimonia can shape their pursuits otherwise. We can see some examples of this in the way Annas and Russell themselves describe the virtuous person in candid moments. Consider what Annas says about bravery:

The properly brave person grasps how acting bravely, while done for its own sake, *also forms part of her overall good in her life* as a whole. ... Bravery involves a correct judgement of the good achieved *in one's life* by acting bravely, rather than in a cowardly way. (Annas, 1993, p. 75, my emphasis)

Notice the reference to good achieved *in one's life*, not the good, period. This is what a coherent eudaimonist should indeed say. And I'm not persuaded that this is consistent with being genuinely virtuous or acting for the sake of others. While something *like* courage may promote the realization of a good life for me considered as a whole and associated positive feelings, the real thing won't fit in the

picture, even if self-interest is not part of the virtuous agent's deliberation. If virtuous activity is part of our final end, then our final end isn't our own happiness.

Fortunately, this dilemma has a simple resolution that allows us to accept most of Aristotle's insights. We should say that virtuous agents pursue the good of others for the sake of leading a good life or the most choiceworthy or successful life, all things considered, and not for the sake of their own happiness, however conceived. That is to say, we should reject the job description that eudaimonists post as far too broad to get at *happiness* rather than something like *the good life* or *the choiceworthy life*. With that, the identification of true happiness with virtuous activity loses its basic rationale.

3. SENTIMENTALISM ABOUT HAPPINESS

I've just argued that being the final end for the sake of which we pursue all other things is not the right job description for *happiness*. Instead, we should begin from a job description that is defined by platitudes like the following: happiness has an experiential quality in that it feels good to be happy; both experienced and expected happiness and unhappiness explain and justify people's life choices; false beliefs can make you happy; different things make different people happy; we can fail to realize that we're happy or unhappy (for discussion, see Kauppinen, Manuscript).

I'm not going to survey the many different accounts of happiness as a psychological state here, but I will jump right into the view that in my opinion best explains such platitudes, Dan Haybron's (2008) *emotional condition* account. According to it, happiness consists in being disposed to have a favourable emotional response to our lives. Such emotional response has three aspects (2008, p. 111–121): *Attunement* consists in emotional states or moods like tranquility, confidence, and imperturbability. When one is attuned to one's environment, one is relaxed and open, "at home" in the world. *Engagement* is manifest in emotions like exuberance and flow, which motivate energetic pursuit of goals. Finally, *endorsement* consists in feelings like joy and cheerfulness, which motivate broadening and building on the things one's life contains. For Haybron, it is above all attunement and engagement that are states of the self and thus crucial to happiness. They are central in the sense that they dispose us to have other affects and moods—in Haybron's slogan, they are productive, persistent, pervasive (permeating the whole of consciousness and setting its tone), and profound (felt as deep) (2008, p. 130–31). In short, then, being happy is psychically affirming one's life in terms of positive central affective states in the three dimensions, as well as having a propensity for positive moods.

Haybron's account explains nicely why happiness feels good (because positive moods and emotions do so), while we can nevertheless be mistaken about how happy we are (because moods, in particular, are in the background of our consciousness and are thus not self-intimating, and our dispositions are not trans-

parent to us). It also makes sense of the explanatory and justificatory heft of happiness in virtue of emphasizing the centrality of many happiness-constituting states to the self, and de-emphasizing the role of peripheral states like sensory pleasures. As a psychological account, it's also compatible with the modern platitude that we can be happy because of false beliefs.

Nevertheless, there are three challenges that can reasonably be raised against Haybron's account. The first is *explanatory depth*: why does happiness consist in just these emotions and moods and not others? The second is what I'll call the *weighting problem*: how do the different dispositions, moods, and emotions combine to determine someone's level of happiness? What determines how much weight each of them has? And, finally, there is the *experientiality question*: how can our dispositions, which are themselves unexperienced, make a difference to our happiness? To see why this last issue is a problem, consider that it's possible (in principle) for two people to have different emotional dispositions and propensities, yet actually experience exactly the same affective states for their whole lives due to, let's say, lucky circumstances in the case of one of them. It seems clear to me that the two people with identical affective experiences over a lifetime must be equally happy, even if only one is *disposed* to be happy. It is, of course, true that one of them has a *happier disposition* and can thus be said to be a *happier person*. And it's typically good for you to have a happy disposition—but only instrumentally good, since it means you're more likely to *be happy*. But all of that is compatible with happiness itself being entirely nondispositional.

My own version of an emotional-condition view, which I develop more fully elsewhere (Kauppinen, Manuscript), is constructed to avoid these problems while building on Haybron's insights. Let's begin with explanation. What is in common to experiences that determine an individual's level of happiness or unhappiness is that the *phenomenal character of each has a valence*, negative or positive, so that they feel good or bad.¹ The default position must surely be that all phenomenal states with a valence contribute to (un)happiness, because they contribute to experiential quality. But what about Haybron's objection that superficial pleasures and pains, like the pleasure of eating a cracker, don't make us any happier or unhappier? On my view, this is a question best answered at the second stage of the theory, which deals with the weighting of valence-bearing experiences in the accounting of one's level of happiness. It is important to note at this point that Haybron is right in emphasizing that not everything we experience at a moment is at the foreground of our attention. A pleasure or a pain can be just the tip of the iceberg. Beneath the water, shaping our experience while easily slipping our attention, lie our emotions and moods. Someone who is sad and someone who is happy may laugh as much, but the totality of their experience will nevertheless be different. Our total affective state at any moment is likely to be complex, nontransparent, and hard to articulate (cf. Bramble, 2016).

How much *weight* do these various valence-bearing experiences have for someone's happiness, then? As Fred Feldman argues, we should start with an account of happiness at a time, or *momentary happiness*, since we can easily define total and average happiness over time in terms of momentary happiness (2010, p. 118–123), so I'll proceed accordingly. The first thing to note is that valence is a matter of degree, whether positive or negative. If we represent the intensity of negative valence with negative numbers, we get the following simple formula for what I'll call *momentary affect balance*:

Momentary affect balance at t is the sum of the intensities of all of S 's valence-bearing phenomenal states at t divided by their number.

If at the moment you're bored with reading the paper to degree -3 , feel self-esteem as a result of comparing yourself to the present author to degree $+7$, and enjoy sitting on the sunny deck to degree $+5$, your momentary affect balance is $-3+7+5/3 = 3$. I think the average rather than the sum better captures the idea of affect balance—otherwise feeling positively about five things would amount to a better total affective state than feeling equally positively about three things.

Momentary affect balance, however, is not the same thing as happiness. For one thing, it is arguably *too volatile*: not every new feeling should make a noticeable difference to our level of happiness. Related to this, it lacks the heft that happiness can have: it's not the kind of thing that explains and justifies life choices. It is also *indiscriminate*: all feelings count equally, regardless of how important to us their objects are. This observation gives rise to a central modification: an affective state's impact on happiness depends not only on its intensity, but also on the *subjective importance* of its object—roughly, the extent to which we value it. For example, the joy you feel when your child recovers from an illness might be as intense as the joy you feel when your favourite team wins a game, but, on the assumption that you value your child's health more than your team's success, the former will make a bigger difference to your happiness.

But why is that, if the affective state feels just as good or bad? I think Haybron is on the right track here in emphasizing the link between happiness and the *self*: the states that count the most for happiness are states of the self, and not just peripheral ones. But unlike him, I don't think that centrality is a matter of disposing agents to experience certain affects rather than others (2008, p. 130), or "output-centrality," as Mauro Rossi and Christine Tappolet (forthcoming) put it. This would result in *double-counting* the contribution of an affect. Suppose in situation A my joy at the intensity of $+5$ gives rise to a $+3$ boost in self-confidence, while in B I experience the same joy, but it doesn't dispose me to feel further positive states. Joy in A is then by definition more output-central than joy in B. But I don't think we should say that the confidence-productive joy in A in itself contributes more to my happiness than the equal joy in B. Of course, it does *causally* contribute more ($+3$ more, to be specific), but it would be double-counting if the *joy* itself had extra weight *in addition to* its own intensity and causal contribution (say, it would contribute $+6$ to my happiness level instead of $+5$).

So, instead, I start from the notion that it is our *values* (or *valuings*) and *concerns* that define our perspective or “self.” By “valuing” an object or person or cause X or a state of affairs Y, I mean, as is fairly standard these days, a complex disposition to pay attention to X or Y, to be pleased when one believes things are going well for X and sad when they’re going badly (or when Y is realized or not), to want things to go well for X (or for Y to be realized), to judge that there is reason for us to have such responses to the fortunes of X or to Y’s reality, and to give weight to X’s interests or Y’s realization in deliberation (e.g., Tiberius, 2018). A “concern” I take to be something more primitive, an attachment to some X that may come apart from judgments about reasons—for example, we may have self-concern even if we don’t value ourselves. Such concern entails motivational and emotional responses to apparent threats to X and to their absence. Both values and concerns belong to the broader class of *sentiments* (Tappolet, 2016). Even among our sentiments, some are more important to who we are than others. We typically have concern for ourselves and our loved ones, and consequently we value our and their honour and health in a different way than we value our favourite television shows, for example. Affects that manifest such core concerns and values, such as worry for a child having trouble at school, can consequently be said to be states of our selves and carry a lot of weight in determining our level of happiness—the child’s troubles *get to us*, as Haybron (2008) says.

Taking this into account, we can define *significance-adjusted momentary affect balance*:

The significance-adjusted intensity of a phenomenal state for S is the product of the intensity of its valence and the subjective importance of its object for S. Significance-adjusted momentary affect balance at *t* is the sum of the significance-adjusted intensities of all of S’s valence-bearing phenomenal states at *t* divided by their number.

(Presumably subjective importance here should range from 0 to 1.) Significance-adjusted momentary-affect balance is, I claim, what we could call “momentary happiness.” The pleasure of eating a cracker doesn’t manifest self-concern or core values—after all, you could have the same pleasure even if your values were different. Of course, there may be people who really value little moments of pleasure in the midst of a busy day and consequently feel joy and delight when they encounter a nice cracker—but then it isn’t implausible that they are a little bit happier for a moment. In contrast, the joy, pride, and self-esteem boost that result from finally getting that big promotion are typically manifestations of valuing professional success and will consequently carry a lot of weight in the happiness of those invested in it.

Significance-adjusted momentary affect balance may still seem too lightweight to capture the deliberative role of happiness—after all, we aren’t, and shouldn’t be, too concerned about momentary happiness when making life choices. But, in my view, as in Feldman’s, we can compare which of two *lives* or *portions of*

lives are happier by adding up or averaging momentary happiness at each relevant moment. And these results—say, how much total happiness I’m going to have in my life going forward—are certainly normatively and explanatorily weighty.

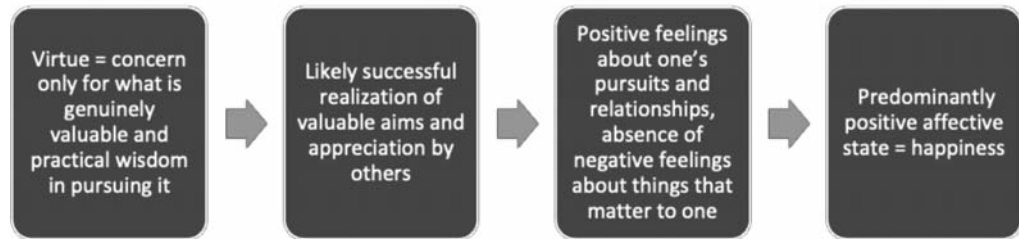
Call the view that identifies happiness with a sufficiently positive sum of significance-adjusted momentary-affect balances over time *sentimentalism about happiness*, or just “sentimentalism” for short (obviously, not to be confused with sentimentalism about moral judgment or metaphysics²). Given typical human concerns, sentimentalism explains many of Haybron’s apt observations about happiness and our emotional condition. Self-concern and concern for loved ones are core sentiments for just about everyone. They manifest themselves in anxiety and depression when our horizon of expectations is threatening or unpromising, and in moods of serenity, tranquility, and optimism when it isn’t. They are also manifest in feelings of self-worth and self-respect when we value ourselves, and in the dreadful feelings of worthlessness otherwise. Through such moods and emotions, we construe our environment as secure or threatening and ourselves as worthy or unworthy of love. According to sentimentalism, these moods and emotions will consequently have a lot of weight in determining our level of happiness—though they rarely take centre stage in our consciousness.

There is more variety in people’s core values, in the importance they give to career, family, causes, hobbies, and other projects. But since the central function of values is to guide our pursuits, emotions that are associated with exercising our agency, such as interest, flow, sense of purpose, and pride (or, conversely, disinterest, feelings of pointlessness, guilt, and shame) will typically have considerable weight in our happiness level in virtue of manifesting core values. We can, of course, value other things, such as the survival of coral reefs or national self-determination, making their fate subjectively important to us and increasing the weight of emotions related to them, including attitudinal pleasures and displeasures (Feldman, 2004). But typically, when it comes to core values in particular, we do tend to do what we can to promote their realization rather than just watch what happens from the sidelines.

Finally, merely sensory pleasure and pains seem to be quite independent of our sentiments, and thus they will as such have little weight in determining our happiness level, even if they’re experiences that have a strong valence. To be sure, it could be said that we have a concern for the felt quality of our experience as such, and thus we will typically have second-order responses to pleasant or painful experiences—I’m attitudinally displeased to be experiencing physical pain, to use Feldman’s language. But here it is those second-order responses that carry weight more significantly in our happiness or unhappiness.

4. WHY THE VIRTUOUS ARE LIKELY TO FEEL GOOD

What does a sentimentalist account of happiness mean for the place of virtue? Clearly, virtuous activity won't be constitutive of happiness, as happiness is not an activity but a total experiential quality. But that doesn't mean that the nature of our activities doesn't make a crucial difference to our affective condition—as I said in the introduction, there's good reason to think virtuous outlook and activities are likely to make us happier than their opposites, as Virtue Conduciveness says. I suggested the following mechanism for this:



Let's now investigate this mechanism in more detail. Recall from section 1 that if we're virtuous, we're good at recognizing what our objectively best realistically achievable options are and at figuring out how to bring them about in an ethically sustainable fashion, because we've been lucky enough to have been habituated into ways of thinking that avoid common human defects like selfishness, shortsightedness, and various biases. Consider what this means for virtuous people's choices. I think that we often end up underselling virtue by focusing on responses to individual situations. After all, practical wisdom isn't most conspicuously exercised in the context of deciding whether to hold a particular position in the face of enemy attack, but rather in the context of choosing whether to go to war in the first place. So, we should think about how virtuous people make more impactful choices in their lives, such as deciding which career to pursue or which project to take on. In these contexts, too, human beings are easily led astray by short-term thinking, selfish focus, or superficial attractions. The practically wise will brush aside such distractions and consider what someone with their particular abilities and in their position can do.

Consequently, the virtuous will be concerned about serving justice and helping those whose flourishing is hindered by their circumstances, and will aim to increase external resources that can be put to good use for their friends and others. So, we can expect them to become pediatricians rather than portfolio managers, public defenders rather than corporate litigators. If they have suitable talent and opportunity, they will conduct research into malaria prevention rather than perform cosmetic surgery, or engineer environment-friendly dams rather than disposable toys. These choices will often mean foregoing money, fame, and comfort, but the virtuous aren't pained by such losses, as long as these don't stand in the way of realizing their aims. Their hearts are in what they do—that is, they emotionally endorse the choices that they make, and don't regret what they've given up or haven't had a chance to pursue. Since they have practical

wisdom, the aims they've selected are realistically achievable and their pursuit of them is intelligent, so they are likely to be successful at realizing their aims, whether these are *prospective* ones that can be realized once and for all or *reflexive* ones of living up to ongoing commitments to, say, family or community (Setiya, 2014; Kauppinen, 2021b). They're also likelier than the nonvirtuous to enjoy good relationships with trustworthy others, because, rather than "takers," they are the kind of "givers" that most people like to be around.

So far, I've described the first two steps of the above diagram. The next and crucial step is that given the kind of values, aims, pursuits, and outcomes that the virtuous are likely to have, they are very likely to enjoy positive-valence-bearing experiences with regard to subjectively important matters. To begin with the *attunement* aspect of their emotional response, we see that since they're likely to be loved and respected by others and to enjoy good friendships (in part because they are more likely to recognize those who lack virtue and avoid them), they're consequently likely to have positive feelings of self-worth. In ordinary circumstances, they're less likely to be anxious or depressed than others, since they don't put as much weight on their own good as others do on theirs. (Evidently, you can find yourself surrounded by vague threats and stuck without appealing prospects even if you're virtuous, so virtue isn't a *guarantee* that negative moods are absent.) But probably the biggest advantage of engaging in virtuous activity is that you're more likely than otherwise to enjoy the pleasures and feelings associated with merited success as well—you'll be able to take pride and joy in what you do. You'll probably enjoy a sense of purpose and healthy self-esteem, and you'll be grateful for the luck you've had in terms of opportunities, upbringing, and external goods. All this means that the virtuous are likely to experience their lives as *meaningful*, which is an important element of happiness (Kauppinen, 2013). Of course, virtuous people may find themselves despised and fail at what they do. If you dedicate yourself to freeing Palestine, your prospects of success are not good. But even failure stings less if you know you're doing the most anyone in your position can do for a good cause. It's also important that the virtuous not be particularly anxious about what other people think of them or how they compare to others, because they don't value success of that sort. Finally, because positive moods and success-related emotions are, as Haybron says, productive of further positive emotions, they make it easy for the virtuous to take joy in little things, like sunshine or a Knopfler song on the radio.

Finally, as I argued in the last section, this overall positive emotional state with respect to things that subjectively matter to the virtuous amounts to happiness. Unlike the eudaimonist picture, sentimentalism considers virtuous activity still to be only a likely *cause* of happiness. So, while I agree with Annas when she says that "our happiness comes at least in part from the way we do or don't actively live our lives, doing something with them or acting in relation to them" (2011, p. 152), I demur when she concludes on this basis that "happiness is at least in part activity" (ibid.)—after all, there's a big difference between being a *source* of happiness and being *constitutive* of happiness. In the debate about

whether virtue is *sufficient* for happiness, sentimentalism takes the Aristotelian side, which says it's not: virtuous people might lack resources, suffer from a chemical imbalance or other disease, and find themselves required to sacrifice their happiness for others. Nor is it necessary. Inside the Experience Machine, one might have the illusion of virtuous activity and be genuinely happy without actually doing anything. This simply follows from the fact that the connection between activity and happiness is causal rather than constitutive.

There are two more interesting ways in which the nonvirtuous, or even the vicious, might gain an equally positive affective condition as the virtuous in favourable circumstances. First, people who lack virtue will sometimes do the right thing and draw enjoyment from it. As a nonvirtuous person, I certainly hope this is the case. But those who lack virtue will not *reliably* track what they have most reason to do in their actions and emotions, and their emotional perspective and judgments may come apart, even if either gets things right. So, sometimes, and perhaps quite a lot of the time, the nonvirtuous will pursue aims that are not worth pursuing and will be likelier to fail at worthy pursuits. I think that in spite of our extraordinary capacity for rationalization, many people do recognize the low worth of their projects at an emotional level. If you choose a job solely on the basis of external rewards, don't be surprised if you feel a lingering dissatisfaction in spite of the stellar sound system in your new Tesla. Many nonvirtuous, of course, succeed in deceiving themselves, in spite of such incipient recognition of worthlessness, and take pride in what they do. This is greatly assisted if they form a community with other nonvirtuous. I have no problem believing that there are happy investment bankers on Wall Street. This goes even for those who are not only nonvirtuous but actually vicious in some way, such as the greedy or the lazy.

I do think, however, that the virtuous have an advantage here. Their positive emotions are actually fitting and reliably so, which makes them *robust* in a way that the positive emotions of the nonvirtuous aren't. The virtuous are also *less vulnerable to luck* when it comes to hitting on a genuinely valuable and thus rewarding activity. It's not a matter of chance that their emotional and rational perspectives form a unified whole. Because they don't pursue pointless or evil aims, they don't need to suspend disbelief or deceive themselves or rely on the unpredictable support of bad company. Nor will gains in knowledge or understanding make them unhappy. So, the positive affective condition of the virtuous depends less on sheer luck than that of the nonvirtuous does.

The second way in which the nonvirtuous in the ethical sense can successfully reach happiness isn't subject to such contingencies. There is, after all, genuine achievement in arts, science, and even business. Such achievements can be sources of pride and self-esteem and beget love and admiration, thus predicting a robustly positive affective condition. Notoriously, they are possible even for the mean, dishonest, and selfish, even if they involve a kind of self-transcendence. But observe that this source of happiness is available only for the few who can really pull it off. It takes a real and special talent to manage and enjoy

artistic success, say, without treating others well, although we know it is strictly speaking possible. But then again, in the real world, moody geniuses are rarely paradigms of radiant happiness!

5. SENTIMENTALIST FAMILY DISPUTES

As I've made clear, my account of happiness owes much to Dan Haybron's emotional-condition view in spite of my rejection of some of its main tenets. In recent years, Mauro Rossi and Christine Tappolet (2016; forthcoming) have also developed an affectivist view of happiness in a different direction from Haybron and have used their account to defend a noncontingent causal connection between virtue and happiness. As we'll soon see, it's fair to describe their account of happiness as a form of sentimentalism. So, it may initially seem that they are the People's Front of Judea to my Judean People's Front. But closer examination of the details of their view reveals significant differences.

Let's start with happiness and its relation to well-being. Like me, and unlike Haybron, Rossi and Tappolet want to include all occurrent, valence-bearing affective states as constituents of happiness, including pleasures and pains. Drawing on Haybron, they then argue that intensity and centrality of an affective state are both relevant to its contribution to happiness. Unlike Haybron, however, they distinguish between two ways in which a state can be central: it can be the *cause* of many other affective states (output centrality) or it can *result* from, or be "grounded in," central affective states or sentiments (source centrality). While Rossi and Tappolet are somewhat ambivalent about the importance of output centrality, I unambiguously reject it as a determinant of how much weight an affective state has for happiness. The role of source centrality, however, is a point of agreement between us.

Building on their past work (Tappolet, 2016; Rossi and Tappolet, forthcoming), Rossi and Tappolet emphasize that happiness-constituting affective states are "phenomenal evaluations" that represent their objects as having evaluative properties. Being composed of such states, happiness itself constitutes "a global affective evaluation." Since such affective evaluations can represent the world correctly or incorrectly, it follows that happiness itself can be *fitting* (both in terms of its constituent affects and the underlying affective dispositions)—that is, represent the subject's evaluative situation correctly.³ The bold claim that Rossi and Tappolet next make is that *well-being*, or what is in itself good for a subject, just is *fitting happiness*. This, they argue, avoids the central objections made to earlier experientialist theories of well-being, such as hedonism—after all, while you may be happy in an experience machine (Lin, 2016), your happiness won't be fitting, because things are not actually the way your positive feelings represent them as being.

Could well-being just be fitting happiness? Rossi and Tappolet's view belongs to the class of experiential-objectivist hybrid theories of well-being, such as Shelly Kagan's (2009) view that well-being is enjoying the (objectively) good.

The basic problem with all such views, as I see it, is that they exclude both positive experiences that are not fitting and objective goods that are unexperienced, as well as corresponding bads. On the first point, consider a simple variation of the Experience Machine thought experiment. Bill and Ted are both in the machine. Bill has a lot of positive experiences, all of which are unfitting, while Ted has none. A strict fitting-happiness view entails that they're equally well off, but they're not—while both are deluded, Bill is at least happy, while Ted isn't. A more relaxed fitting-happiness view (similar to one of Fred Feldman's (2004) suggestions) says that while fitting positive feelings are *better* for you than unfitting ones, unfitting ones are also good. Such a view would explain the asymmetry between Bill and Ted, but it would lack the explanatory unity that Rossi and Tappolet claim as an advantage of their view—now we have two kinds of final good, fitting and unfitting happiness. What's more, it's not clear whether there's any rationale for giving more weight to fitting than unfitting happiness that's not *ad hoc*—after all, *as an experience*, unfitting happiness seems to be just as good as fitting happiness.

The other basic problem for experiential-objectivist hybrids is that just as they rule out enjoyment of the nonvaluable, they rule out unenjoyed values. Anything that is in itself good for us without affecting our happiness is a counterexample. Many people believe in posthumous benefits, which we may well not be fittingly happy about (though you could, in theory, correctly anticipate and take pride in posthumous success while you live!). Less controversially, while we live, our goals may be realized or our trust betrayed without our finding out about it. Indeed, such things may never make a difference to our experience. Nevertheless, they seem to make us better or worse off (e.g., Nagel, 1979)—they are the kind of things we'd intrinsically want for those we love for their sake (Darwall, 2002). Indeed, some valence-bearing affects seem to be fitting precisely because they correctly represent things as going nonexperientially well or badly *for us*, which puts the fitting-happiness view in a bind. Suppose I'm relieved to discover I wasn't betrayed after all. Such relief is presumably fitting only if it's bad for me to be betrayed independently of my feelings about betrayal (what precisely I'm relieved about is the fact that I wasn't betrayed, *not* that I was spared the bad feelings that would have arisen upon my knowledge of events had I been betrayed). But if that's the case, some instances of fitting happiness presuppose that things other than fitting happiness can be welfare goods.⁴

Rossi and Tappolet (forthcoming) may think that the above consequences of their experiential-objective hybrid view are worth taking on board, because their view offers explanatory unity that treating happiness and, say, agential success as potentially distinct welfare goods lacks. It is evidently not possible to discuss all the alternative ways of unifying welfare goods here. But just to give a hint of the view I prefer, on the version of perfectionism I'm developing (for an initial statement, see Kauppinen, 2015, p. 218–219), what unifies the things that are in themselves good for us is that they constitute realizations of the overarching aims that are implicit in the exercise of our essential capacities as the kind of well-being subjects that we are. In the adult human case, this means that our

well-being consists in developing and exercising our capacity both for temporally extended rational agency and for positive experience, since we're essentially both extended rational agents and subjects of valence-bearing experiences.⁵ Such perfectionism, I believe, provides the desired explanatory unity while allowing the two dimensions of prudential value to vary independently, thus avoiding the problems of hybrid views while capturing their attractions.

Finally, let me turn to the link between virtue and happiness in Rossi and Tappolet's work. Instead of the Aristotelian picture of virtue, they endorse a more minimal account, according to which a virtue just consists in disposition to experience fitting emotions, which further disposes one to act and judge correctly (2016, p. 116). For reasons noted in the first two sections, I think the Aristotelian view gives a better account of practical reasoning and moral motivation. But set that aside and focus on the noncontingent link between virtue and happiness they propose:

If virtue involves a disposition to experience fitting emotions, then, to the extent that the external circumstances are propitious, virtue disposes the individual to experience positive fitting emotions. (Rossi and Tappolet, 2016, p. 122)

Someone's external circumstances are said to be "propitious" "when the individual is confronted with positive evaluative properties or, more simply, to [i.e., with] positive values" (2016, p. 123). Note that this is a much weaker noncontingent connection between virtue and happiness than my view makes. Basically, what Rossi and Tappolet seem to say is that if you happen to find yourself in good circumstances, then, if you're virtuous, you're also happy. In contrast, what I say is that if you're virtuous, you're likely to *bring about* good circumstances (this is the Aristotelian part) and you're likely not to mind most things that actually cause bad feelings (this is the Stoic part). (No doubt this issue is related to the passivity worry about Rossi and Tappolet's view of well-being.) So the mechanism for the noncontingent link between virtue and sentimentalist happiness that they suggest is very different from mine, which highlights the role of virtuous activity and related emotions.

CONCLUSION

I don't know about you, but I'd like to be happier (and, sure, fitter and more productive, too). I've argued that what this would amount to would be having more positive feelings—and fewer negative feelings—about what I most care about. I've also argued that since virtuous people not only have the right values and concerns but also have the skills to pursue these wisely, they're likely to be successful and to be liked. At the same time, they don't give a damn about things that don't really matter. Consequently, virtue and virtuous activity are likely to be relatively conducive to happiness in comparison to nonvirtue, though this noncontingent tendency is still subject to luck. Moreover, a virtuous person's happiness doesn't depend on self-deception, mistake,

or illusion, and it doesn't require extraordinary talent. It is thus effortless to sustain and, in the real world, more likely to persist than happiness based on unfitting positive emotion. Being convinced of these things, I'd like to be more virtuous for purely selfish, borderline-hedonistic reasons, as well as for ethical ones. Alas, it's easier said than done for an old dog. It's doubtful whether even acting as the virtuous do would improve my character. But, as Aristotle says, there's little else I can do, so I had better go and answer some student emails.⁶

NOTES

- ¹ For my own relational imperativist account of valence, see Kauppinen, 2021a.
- ² For different things one can be a sentimentalist about, see Kauppinen, 2014.
- ³ I have a different account of fit, according to which, roughly, an emotion is fitting when its indicative content is a reason to respond as its imperative content tells us to do (for an early take on this type of view, see Kauppinen, 2014b). This difference doesn't matter for current purposes, so I'll set it aside.
- ⁴ Rossi and Tappolet might say in response that what makes my relief fitting is that my being betrayed is bad *simpliciter*, not bad for me (2021, p. 13). But I think that my relief represents it as being bad for me, however, and also that even if betrayal is, say, morally bad, it's also specifically *bad for me* to be betrayed.
- ⁵ Rossi and Tappolet themselves raise the issue of the role of agency in well-being for their view. For them, agency enters into well-being only contingently, via agency-linked emotions like flow and possible agency-linked valuings (2021, p. 20). You could, in principle, be fittingly happy—and thus do well or flourish—without doing anything, just enjoying the genuinely good things that happen around you. On my perfectionist view, in contrast, there would be something very important missing in your life.
- ⁶ I want to thank people who gave feedback on this paper and the talk that preceded it, including but not limited to Erik Angner, Dan Haybron, Richard Kraut, Stéphane Lemaire, Lilian O'Brien, Mauro Rossi, and Christine Tappolet, and audiences in Dublin, Stockholm, and Clermont-Ferrand. I owe special thanks to Anna Alexandrova, without whose encouragement I would have set the paper aside.

REFERENCES

- Annas, Julia, *The Morality of Happiness*, New York, Oxford University Press, 1993.
- , *Intelligent Virtue*, Oxford, Oxford University Press, 2011.
- Aristotle, *Nicomachean Ethics*, translated and edited by Roger Crisp, Cambridge, Cambridge University Press, 2000.
- Bramble, Ben, “A New Defense of Hedonism about Well-Being”, *Ergo: An Open Access Journal of Philosophy*, vol. 3, no. 4, 2016, p. 85-112.
- Brennan, Tad, *The Stoic Life: Emotions, Duties, and Fate*, Oxford, Oxford University Press, 2005.
- Darwall, Stephen, *Welfare and Rational Care*, Princeton, Princeton University Press, 2002.
- Doris, John M., *Lack of Character: Personality and Moral Behavior*, Cambridge, Cambridge University Press, 2002.
- Feldman, Fred, *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*, Oxford, Clarendon Press, 2004.
- , *What is This Thing Called Happiness?*, Oxford, Oxford University Press, 2010.
- Foot, Philippa, *Virtues and Vices and Other Essays in Moral Philosophy*, Berkeley, University of California Press, 1978.
- , *Natural Goodness*, New York, Oxford University Press, 2001.
- Haybron, Daniel M., *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, Oxford, Oxford University Press, 2008.
- , *The Good Life*, Manuscript.
- Hills, Alison, “Understanding Why”, *Noûs*, vol. 50, no. 4, 2016, p. 661-688.
- Kagan, Shelly, “Well-Being as Enjoying the Good”, *Philosophical Perspectives*, vol. 23, no. 1, 2009, p. 253-272.
- Kahneman, Daniel, *Thinking, Fast and Slow*, New York, Farrar, Straus and Giroux, 2011.
- Kauppinen, Antti, “Meaning and Happiness”, *Philosophical Topics*, vol. 41, no. 1, 2013, p. 161-185.
- , “Moral Sentimentalism”, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2014.
- , “The Narrative Calculus”, *Oxford Studies in Normative Ethics*, vol. 5, 2015, p. 196-220.
- , “Relational Imperativism About Affective Valence”, *Oxford Studies in the Philosophy of Mind*, vol. 1, 2021a, 341-371.

———, “Against Seizing the Day”, *Oxford Studies in Normative Ethics*, vol. 11, 2021b, p. 91-111.

———, “Happiness, Desire, and Emotion”, Manuscript.

Kraut, Richard, “Two Conceptions of Happiness”, *Philosophical Review*, vol. 88, no. 2, 1979, p. 167-197.

Lin, Eden, “How to Use the Experience Machine.” *Utilitas*, vol. 28, no. 3, 2016, p. 314-332.

McDowell, John, “Are Moral Requirements Hypothetical Imperatives?”, *Proceedings of the Aristotelian Society*, vol. 52 (Supplement), 1978, p. 13-29.

McDowell, John, “Virtue and Reason”, *The Monist*, vol. 62, no. 3, 1979, p. 331-350.

Nagel, Thomas, *Mortal Questions*, New York, Cambridge University Press, 1979.

Nussbaum, Martha C., *Anger and Forgiveness: Resentment, Generosity, Justice*, New York, Oxford University Press, 2016.

Phillips, Jonathan, Sven Nyholm, and Shen-yi Liao, “The Good in Happiness”, in Tania Lombrozo, Shaun Nichols and Joshua Knobe (eds.), *Oxford Studies in Experimental Philosophy, Volume 1*, Oxford, Oxford University Press, 2014, p. 253-293.

Rossi, Mauro and Christine Tappolet, “Virtue, Happiness, and Wellbeing”, *The Monist*, vol. 99, no. 2, 2016, p. 112-127.

———, “Well-Being as Fitting Happiness”, in Christopher Howard and Richard Rowland (eds.), *Fittingness*, Oxford, Oxford University Press, forthcoming.

Russell, Daniel C., *Happiness for Humans*, Oxford, Oxford University Press, 2012.

Setiya, Kieran, “The Midlife Crisis”, *Philosophers’ Imprint*, vol. 14, no. 31. 2014.

Smith, Michael, *The Moral Problem*, Oxford, Blackwell, 1994.

Tappolet, Christine, *Emotions, Value, and Agency*, New York, Oxford University Press, 2016.

Thaler, Richard and Sunstein, Cass, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, Yale University Press, 2008.

Tiberius, Valerie, *Well-Being as Value Fulfillment: How We Can Help Each Other to Live Well*, New York, Oxford University Press, 2018.

LES BONNES HUMEURS, LE BONHEUR ET LE BIEN-ÊTRE

SAMUEL LEPINE

UNIVERSITÉ CLERMONT AUVERGNE, LABORATOIRE PHIER

RÉSUMÉ :

Dans cet article, je propose d'analyser le concept de bonheur en termes d'humeurs positives. Je montre que cette analyse constitue une voie moyenne entre l'analyse du bonheur comme émotion et l'analyse du bonheur en termes de propensions émotionnelles. Je soutiens plus particulièrement qu'être heureux consiste à ressentir une humeur positive, quelle que soit cette humeur. Cette analyse est donc réductionniste – être heureux n'est rien d'autre qu'être de bonne humeur – et pluraliste – nos bonnes humeurs, dans toute leur diversité, constituent autant de manières différentes d'être heureux. Afin de soutenir cette approche, je défends la thèse selon laquelle les humeurs sont des attitudes prospectives, consistant à évaluer l'impact affectif possible de notre environnement sur nos états motivationnels tels que nos désirs ou nos préférences. Je m'efforce ensuite d'explorer les relations que le bonheur entretient avec le bien-être à partir de cette analyse du bonheur comme bonne humeur.

ABSTRACT:

In this paper, I propose to analyze the concept of happiness in terms of positive moods. I show that this analysis constitutes a middle way between the analysis of happiness as an emotion, and the analysis of happiness in terms of mood propensities. In particular, I argue that being happy consists in feeling a positive mood, whatever that mood may be. This analysis is therefore reductionist - being happy is nothing more than being in a good mood - and - pluralist - the diversity of our good moods constitutes the many different ways of being happy. In order to defend this approach, I argue that moods are prospective attitudes, consisting in evaluating the possible affective impact of our environment on our motivational states, such as our desires or preferences. I then explore the relationship between happiness and well-being based on this analysis of happiness as a good mood.

INTRODUCTION

Il est difficile de saisir ce que recouvre le concept de bonheur, en dépit du rôle important qu'il semble jouer dans nos vies. On se le représente parfois comme un simple état ressenti, ponctuel, dans lequel nous nous trouvons lorsque nous passons un bon moment. On dit ainsi que l'on a éprouvé du bonheur en passant la journée en compagnie de nos amis, en assistant à un concert, ou en mangeant un plat raffiné. Prenant appui sur cette intuition, certains philosophes ont développé la thèse selon laquelle le bonheur serait avant tout une émotion¹. D'autres résistent cependant à cette idée, en s'appuyant sur l'intuition contraire selon laquelle le bonheur serait, plus qu'un simple épisode émotionnel, un état qui s'inscrit dans la durée. Ils proposent de caractériser le bonheur en faisant plutôt référence à des « propensions » émotionnelles, telles que des dispositions à éprouver des humeurs et des émotions positives sur le long terme². De telles propensions ne seraient pas systématiquement et toujours ressenties, mais cette même propriété permettrait justement de rendre raison du fait que le bonheur n'est pas qu'une émotion plus ou moins fugace. Nos intuitions au sujet du bonheur sont ainsi écartelées entre deux conceptions contradictoires, et les théories philosophiques reflètent cet écartèlement. Mon but, dans cet article, sera de tenter d'analyser le concept de bonheur au-delà de ce conflit d'intuitions, en mettant l'accent sur le concept d'humeur. Cette question revêt une certaine importance, car elle conditionne aussi le type d'état auquel nous devrions nous intéresser lorsque nous essayons de mesurer le bonheur des individus (ce que les psychologues appellent de leur côté le « bien-être subjectif »). Or, le fait est que les humeurs n'ont jusqu'ici reçu que très peu d'attention dans les études empiriques sur le bonheur³, alors qu'elles sont souvent au cœur des enquêtes sur la dépression (Wakefield et Demazeux, 2016).

De fait, aucun auteur ne semble avoir sérieusement envisagé la possibilité que le bonheur soit un état intermédiaire entre le caractère épisodique des émotions et le caractère durable, mais non ressenti, que revêtent les dispositions à éprouver des humeurs et des émotions. La thèse que je défendrai ici est que le bonheur serait pourtant mieux compris comme une humeur positive, et qu'être heureux se réduit à être de bonne humeur, c'est-à-dire à éprouver l'une des humeurs positives que nous sommes susceptibles de ressentir au cours d'une vie humaine. Après tout, lorsque nous déclarons que nous sommes heureux, c'est à notre humeur générale que nous faisons référence. L'indice sur lequel nous nous basons, lorsque l'on nous demande comment nous allons, est notre humeur également. En dépit de son caractère attrayant, cette approche n'a pourtant jamais été explorée pour elle-même. Dans cet article, je me propose donc de comprendre le concept de bonheur en m'arrêtant sur l'analyse du concept d'humeur et en soutenant que le bonheur n'est rien d'autre qu'une humeur positive, quelle que soit la nature de cette humeur (enjouement, excitation, calme, etc.). Cette thèse est réductionniste, en ce qu'elle soutient qu'il n'y a rien de plus dans le bonheur qu'une bonne humeur. De fait, il semble conceptuellement impossible d'être à la fois de bonne humeur et malheureux⁴. Mais c'est aussi, comme nous le verrons, une thèse pluraliste en ce qu'elle admet en même temps qu'il y a de nombreuses manières d'être heureux.

Afin de défendre cette thèse, je commence par passer en revue les théories émotionnelles du bonheur et je montre ensuite que la plupart des propriétés que l'on attribue au bonheur conçu comme émotion sont en réalité mieux comprises lorsque l'on comprend le bonheur comme bonne humeur. Je soutiens également qu'il n'est pas nécessaire de concevoir le bonheur en termes de propensions à éprouver des humeurs positives. À l'encontre de ces deux approches, je propose, en m'inspirant des analyses de Nesse (1991), de concevoir les bonnes humeurs comme des évaluations positives de la « propiciosité » (*propitiousness*) de notre situation personnelle, c'est-à-dire de la probabilité que celle-ci soit propice à la satisfaction de nos états motivationnels tels que nos désirs ou nos préférences. Je défends plus précisément la thèse selon laquelle les humeurs sont des attitudes prospectives, qui consistent à évaluer l'impact affectif possible de notre environnement sur nos motivations⁵. Pour finir, je m'efforce d'explorer les relations que le bonheur entretient avec le bien-être à partir de cette analyse du bonheur comme bonne humeur. Je soutiens que les bonnes humeurs ont une valeur prudentielle lorsqu'elles sont appropriées, c'est-à-dire lorsqu'elles évaluent correctement la propiciosité des situations auxquelles nous sommes confrontés.

1. LE BONHEUR COMME ÉMOTION

Les auteurs qui soutiennent que le bonheur est une émotion insistent sur le caractère épisodique du bonheur. Il semble en effet que l'on puisse être heureux de réussir un examen, de regarder un bon film, de jouer au basket, etc. L'émotion de bonheur aurait ainsi une intentionnalité évaluative, au sens où, comme toute émotion, elle porterait sur un objet intentionnel particulier (un succès, un film, une activité, etc.), qu'elle évalue positivement. L'émotion de bonheur aurait aussi, comme toute émotion, un caractère phénoménal, au sens où elle serait nécessairement ressentie et se caractériserait par certaines sensations. Cette émotion possède également une valence positive, dans la mesure où il est en principe plaisant de l'éprouver. La valence, en ce sens, désigne la tonalité hédonique d'un état affectif (émotion, humeur, etc.), qui peut être positive lorsque ce dernier est plaisant et négative lorsqu'il est déplaisant.

On peut toutefois envisager deux manières au moins de comprendre cette caractérisation du bonheur comme émotion. Selon une première approche, réductionniste, le bonheur se réduit en réalité à d'autres émotions telles que la joie, l'enthousiasme, l'extase, la gaieté, ou encore l'amusement. Selon l'approche réductionniste, être heureux reviendrait à éprouver l'une ou l'autre de ces émotions positives. De fait, dans les différents exemples mentionnés plus haut, le bonheur semble précisément être pris comme un synonyme de ces émotions : on est joyeux de réussir son examen, enthousiasmé de jouer au basket, etc. Mais les auteurs qui affirment que le bonheur est une émotion entendent généralement défendre une thèse non réductionniste, notamment parce que l'on pourrait très bien éprouver ces émotions sans véritablement être heureux. La question qui se pose alors est la suivante : en quoi le bonheur comme émotion pourrait-il bien se distinguer des autres émotions positives mentionnées dans les exem-

ples précédents ? Il existe au moins deux propositions dans la littérature actuelle. La première consiste à insister sur les jugements évaluatifs caractéristiques de l'émotion de bonheur, la seconde à mettre l'accent sur l'objet spécifique de cette émotion. Dans ce qui suit, j'examinerai ces deux approches afin d'en souligner les limites et de montrer qu'une analyse du bonheur comme bonne humeur est préférable à celles-ci.

Afin de défendre la première approche, Goldman (2017) soutient que le bonheur est une émotion ponctuelle par opposition à un état de longue durée. Il caractérise les émotions comme des épisodes psychologiques incluant à la fois un jugement et des sensations. L'émotion de bonheur consisterait selon lui pour une part importante⁶ en un jugement de satisfaction portant sur notre vie, au moment où nous prononçons ce jugement, mais aussi en un genre de « sensation pure » (2017, p. 10), qui aurait la particularité d'être dotée d'une teinte phénoménale propre, sans être localisée dans une partie précise de notre corps (et c'est en cela qu'elle serait « pure ») : c'est ce que l'on ressent par exemple lorsque l'on reçoit une bonne nouvelle. Être heureux reviendrait ainsi à juger que notre vie va bien au moment où nous éprouvons cette émotion, et l'émotion de bonheur aurait pour objet intentionnel notre vie à un certain moment, ou éventuellement un aspect plus étendu de celle-ci. Le jugement caractéristique de l'émotion de bonheur, selon Goldman, est donc une évaluation positive de notre vie ou de certains aspects de celle-ci, et ce jugement peut être implicite ou explicite. Le bonheur peut ainsi être inauthentique si le jugement caractéristique de notre émotion évalue mal la situation, par exemple parce que j'éprouve du bonheur en pensant à la réussite de ma vie de couple, alors que ma compagne ne m'aime plus sans que je le sache.

Cette première théorie émotionnelle du bonheur s'efforce ainsi de rendre compte du caractère à la fois ressenti et évaluatif du bonheur. Si Goldman semble soutenir une conception relativement intellectualiste des émotions (notamment en les assimilant en partie à des jugements), il est clair que son approche est compatible avec différentes théories des émotions. On pourrait très bien imaginer par exemple que le bonheur comme évaluation positive de notre vie soit une forme de perception évaluative (comme le soutient par exemple Tappolet, 2016, au sujet des émotions en général). Pour autant, il faut bien reconnaître que Goldman n'est pas clair lorsqu'il aborde la nature évaluative du bonheur comme émotion. Selon lui, nous l'avons vu, je suis heureux si je juge que ma vie va bien et que je ressens certaines sensations caractéristiques de l'émotion de bonheur. Mais dans ce cas, est-ce que mon émotion n'est pas juste une forme de réjouissance ou de joie ? Dire que j'éprouve du bonheur en évaluant positivement un aspect de ma vie, n'est-ce pas simplement une métaphore pour dire que je me réjouis de cet aspect de ma vie, ou qu'il me remplit de joie ? Goldman ne fournit pas réellement d'argument qui devrait nous permettre d'écarter une interprétation réductionniste de ce genre. Et si tel est le cas, alors on voit difficilement ce qu'aurait de spécial le bonheur en tant qu'émotion. Être heureux ou éprouver du bonheur reviendrait simplement à dire que l'on se réjouit d'un aspect de notre vie. Peut-être est-ce en fait la thèse que Goldman entend soutenir, mais dans ce cas, on

peut s'étonner qu'il ne l'affirme pas plus clairement, et il eut été souhaitable qu'il clarifie alors davantage la nature évaluative du bonheur en tant qu'émotion.

Une autre option consiste à essayer de cerner plus précisément ce que certains appelleraient l'objet formel du bonheur. L'objet formel d'une émotion, en effet, est censé nous permettre de préciser le type d'évaluation qui caractérise spécifiquement cette émotion (Deonna et Teroni, 2012). On peut dire ainsi que le danger est l'objet formel de la peur dans la mesure où un épisode de peur consiste en l'évaluation d'un objet comme étant dangereux. Différents épisodes de peur peuvent ainsi porter sur différents objets particuliers (une voiture qui nous fonce dessus, l'apparition d'un nouveau virus, la montée de l'extrême droite, etc.), mais tous ces objets ont en commun de représenter un danger pour nous. De même, la fierté peut avoir différents objets particuliers (un succès lors d'un concours, un plat que l'on a préparé, un jardin que l'on entretient, etc.), mais ce qui caractérise la fierté est probablement d'évaluer ces objets comme des formes de réussites personnelles. L'objet formel d'une émotion, en ce sens, désigne une valeur négative (le danger) ou positive (la réussite) qui permet de déterminer l'évaluation caractéristique de cette émotion.

En suivant une stratégie de ce genre, Mulligan (2016) souligne que le bonheur comme émotion peut certes porter sur différents objets particuliers, tels qu'une victoire ou un moment passé entre amis, mais que ces différents objets particuliers ont en commun d'être des biens ou des situations que nous sommes chanceux d'avoir. L'objet formel du bonheur serait donc la bonne fortune à laquelle un individu doit son épanouissement⁷. Afin de défendre cette thèse, Mulligan soutient que le bonheur ne porterait pas sur un objet ou sur une personne en particulier – contrairement à de nombreuses émotions – mais sur le fait qu'il soit bon que certaines valeurs épaisses soient présentes. Son argument repose sur « le principe de l'ascension axiologique », qui stipule que si un fait exemplifie une valeur épaisse alors il exemplifie nécessairement une valeur fine. Un concept de valeur épais (*thick ethical concept*), en effet, est un concept dont le contenu est à la fois descriptif et évaluatif (comme le « dégoûtant »), là où un concept de valeur fin (*thin*) est un concept plus général et purement évaluatif (comme « mauvais »). Le principe de l'ascension axiologique semble donc évident, puisque le concept de « dégoûtant » implique manifestement celui de « mauvais », tout comme le concept de « réjouissant » implique le concept de « bon ».

Ainsi, si un fait exemplifie de la justice, ou du plaisir, alors il est bon que ce soit le cas, dans la mesure où ces valeurs épaisses impliquent des valeurs fines. Le bonheur aurait ainsi pour objets particuliers différentes formes de bontés : le fait de rencontrer des succès, d'être admirable, de ressentir de la gratitude, d'avoir des enfants aimables, etc. Pour le dire autrement, les objets particuliers du bonheur seraient des valeurs positives exemplifiées par certains faits, mais l'objet formel du bonheur désignerait le fait que l'on soit chanceux de rencontrer des succès, d'être admirable, etc., bref, qu'il en soit ainsi. Cette idée capture le fait trivial que, dans de nombreuses langues, le bonheur implique étymologi-

quement l'idée de chance, mais elle renvoie aussi plus fondamentalement au fait que notre situation est globalement bonne (Mulligan, 2016, p. 139)⁸. Être heureux ne supposerait pas simplement de voir certains de nos désirs satisfaits, mais d'être en présence de valeurs qui témoignent de notre chance, telles que la justice, l'amitié, ou même la possession de certaines capacités.

L'un des attrait des objets formels est qu'ils permettent de rendre intelligibles nos émotions et d'évaluer leur correction (Teroni, 2007). Dans le cas de la théorie émotionnelle du bonheur de Mulligan, cela revient à dire, d'une part, qu'il serait en effet absurde d'être heureux à propos de situations dans lesquelles nous sommes malchanceux, au sens où une telle réaction serait inintelligible : on ne peut pas se réjouir en principe de manquer de chance. Et cela revient à dire, d'autre part, qu'il serait correct d'être heureux si et seulement si l'on a de la chance, si le monde nous est globalement favorable. Il serait par conséquent incorrect d'être heureux si l'on se réjouit de la valeur positive de capacités, de relations ou de situations qui ne sont pas des exemples de notre bonne chance. Comme le souligne Mulligan, l'objet de notre émotion de bonheur peut cacher de grandes illusions, et je peux être doté de grandes capacités pour un sport qui n'est pas pratiqué là où je réside.

Cette thèse possède au moins deux attrait. Premièrement, elle rend compte du fait que le bonheur semble souvent au-delà de notre contrôle. Il s'apparente en cela à un genre d'effet secondaire que l'on ne peut obtenir délibérément, ou encore à un « cadeau » comme on le dit parfois. Deuxièmement, elle essaie de tisser un lien entre nos différentes émotions positives et le bonheur lui-même. Le bonheur, en effet, est compris ici comme une émotion réflexive, qui vient du fait que nous prenons du recul sur notre propre vie, voire encore comme une forme de méta-émotion au sujet de nos émotions ordinaires. Ce faisant, l'approche de Mulligan permet de mettre le doigt sur le fait que le bonheur est généralement tourné vers des situations plus générales que de simples objets, et notamment sur notre bonne fortune, au sens où le bonheur serait une manière d'évaluer le fait que les choses vont « globalement » bien pour nous.

D'un autre côté, cette approche semble rencontrer un certain nombre de difficultés. Il faut noter en premier lieu qu'il n'est pas évident de déterminer à partir de quand exactement la possession d'un bien ou d'une capacité est due à la chance. Les situations qui doivent être considérées comme « chanceuses » sont également loin d'être évidentes, étant donné que toute situation apparemment positive peut s'avérer en réalité défavorable. En l'état, le critère de la chance semble donc excessivement relatif. La deuxième difficulté que rencontre cette thèse est similaire à celle que rencontrait déjà la thèse de Goldman. En effet, on ne voit pas vraiment en quoi l'émotion de bonheur dont il est ici question est censée être différente du simple fait de se réjouir qu'une valeur positive soit présente. La différence semble purement lexicale une fois encore. Une réponse possible pourrait être que l'objet formel du bonheur n'est pas le même que celui de l'émotion de réjouissance, puisque la première ne porte pas sur le réjouissant mais sur la bonne fortune. Mais cette réponse court le risque d'être exagérément

intellectualiste : lorsque nous sommes heureux, notre émotion ne semble pas porter sur le fait que nous avons de la chance. Il peut arriver qu'en méditant sur ce genre de considérations, nous nous sentions effectivement chanceux, mais il n'est pas nécessaire de réfléchir sur la valeur positive d'une situation pour être heureux.

Une troisième difficulté, enfin, que l'on retrouve là encore aussi bien dans l'approche de Goldman que dans celle de Mulligan, tient au fait que le bonheur soit réduit à un épisode temporel aussi bref qu'une émotion. Bien sûr, on peut admettre que le concept de bonheur est polysémique et qu'il peut désigner différents types d'états. Sumner (1996) distingue ainsi quatre sens possibles du terme « heureux » (*happy*), au premier rang desquels vient le fait d'éprouver une émotion et d'être heureux (*being happy with*) au sujet de quelque chose. C'est précisément à cet aspect que fait référence la théorie émotionnelle du bonheur. Mais si celle-ci ne nous donne pas de raison satisfaisante d'admettre que l'émotion de bonheur est différente du fait de se réjouir, alors il n'est pas clair que l'on puisse considérer cette acception comme une caractérisation authentique du bonheur. Sumner lui-même l'écarte en affirmant qu'« être heureux de quelque chose est à peu près équivalent au fait d'être satisfait ou content au sujet de cette chose, c'est-à-dire trouver qu'elle est raisonnablement à la hauteur de certaines attentes » (Sumner, 1996, p. 143-144). Inversement, on peut « se sentir heureux » (*feeling happy*), au sens où l'on éprouverait de façon plus généralisée un ressenti positif qui ne serait pas dirigé sur un objet particulier. Ce deuxième sens est plus proche d'une humeur susceptible de colorer l'ensemble de notre regard sur notre propre vie, comme le note Sumner. On peut se sentir plus ou moins heureux, en ce sens, sur un continuum qui va de l'exaltation à la dépression.

C'est bien cette deuxième signification qui semble embrasser le sens ordinaire du bonheur, lequel a une forme plus diffuse et plus longue que le simple fait d'éprouver de la joie ou le fait de se réjouir, qui sont des épisodes émotionnels temporellement réduits. Les humeurs, au contraire, désignent des états affectifs temporellement plus longs que les émotions et plus diffus que ces dernières. À côté de ces deux significations de « heureux », Sumner distingue encore le fait d'avoir une personnalité heureuse, en ce que l'on est disposé, par un tempérament enjoué, enthousiaste, ou optimiste, à éprouver des humeurs positives et ainsi à être généralement une personne heureuse. Ce troisième sens ne renvoie donc qu'indirectement au bonheur et n'est pas vraiment pertinent ici. Enfin, Sumner distingue le fait d'avoir une vie heureuse (*having a happy life*), qui désignerait le fait d'avoir une attitude positive envers sa propre vie. Cette attitude serait à la fois affective et cognitive, en ce qu'elle renvoie à la fois au fait de se sentir satisfait par sa vie et au fait d'évaluer positivement sa vie, que ce soit sous l'un de ses aspects au moins, ou bien dans sa globalité. Le bonheur, ainsi compris, consisterait en « un jugement que, dans l'ensemble et en prenant tout en compte, votre vie va bien pour vous » (Sumner, 1996, p. 146). Le bonheur consisterait alors en une forme d'endossement ou d'acceptation de nos conditions de vie.

On peut toutefois se demander si cette proposition n'est pas, une fois encore, trop intellectualiste. Sumner reconnaît lui-même que la dimension cognitive de cette théorie exclut que les enfants en bas âge et les animaux puissent être heureux en ce sens, bien qu'ils puissent l'être d'un point de vue affectif (Sumner, 1996, p. 147). Il n'est peut-être pas nécessaire d'aller aussi loin. Dans la prochaine section, j'aimerais proposer l'idée que la deuxième signification de « heureux » que dégage Sumner est suffisante pour comprendre ce que recouvre le concept de bonheur : le bonheur peut en effet se caractériser simplement comme une « bonne humeur ». Cette proposition nécessite simplement une analyse plus approfondie du concept d'humeur.

2. LE BONHEUR COMME HUMEUR POSITIVE

Peut-on rendre compte de façon satisfaisante du fait que le bonheur est une humeur positive ? Il est nécessaire pour accomplir cette tâche de recenser en premier lieu les caractéristiques principales des humeurs afin de voir dans quelle mesure exactement elles peuvent apporter un éclairage satisfaisant sur le concept de bonheur. Dans cette section, je m'efforce de motiver l'idée selon laquelle le bonheur pourrait n'être qu'une humeur positive.

Il faut noter en premier lieu que nos humeurs sont liées aux émotions que nous éprouvons le plus souvent et qu'elles tirent généralement de ces dernières leur propre valence positive ou négative. Nous sommes de mauvaise humeur après avoir éprouvé des émotions négatives et, inversement, nous sommes de bonne humeur lorsque nous éprouvons davantage d'émotions positives, et le fait d'être heureux semble coïncider précisément avec ce genre d'état. En ce sens, le bonheur peut désigner l'humeur positive qui résulte elle-même d'émotions positives. Une telle hypothèse semble beaucoup moins coûteuse que l'affirmation selon laquelle le bonheur est une émotion réflexive portant sur notre vie. De façon beaucoup plus simple, les humeurs héritent des évaluations effectuées par nos émotions au sujet des situations que nous rencontrons et qu'elles généralisent. Si la colère consiste à évaluer une situation comme offensante, une humeur noire consiste à percevoir toute situation comme étant susceptible d'être offensante. De ce point de vue, les humeurs et les émotions se situent sur un continuum évaluatif (Siemer, 2009).

Il y a donc à première vue une continuité évidente entre les émotions et les humeurs. Il faut néanmoins se garder de ne voir dans les humeurs que des émotions plus générales (comme le soutient notamment DeLancey, 2006). Il existe en effet plusieurs différences importantes entre les émotions et les humeurs. La première est que les humeurs se déploient graduellement dans le temps, tandis que les émotions au contraire ont en général un début assez facilement identifiable (en ce qu'il est lié à un événement précis) et une fin que l'on peut également essayer de cerner dans une certaine mesure (Larsen, 2000). De ce point de vue, les humeurs ont un caractère plus diffus que les émotions, tout en étant aussi parfois plus envahissantes (Tappolet, 2018). Elles peuvent s'emparer progressivement de nous, colorer l'ensemble de nos pensées, puis disparaître également de façon imperceptible.

Une autre différence est que les émotions sont tournées vers le monde extérieur alors que les humeurs semblent parfois davantage tournées vers nous-mêmes : elles nous informent à propos de nos ressources pour répondre aux menaces et aux défis que pose notre environnement (Morris, 1992). Se sentir d'humeur déprimée, c'est sentir aussi que l'on n'est pas en mesure de faire face aux difficultés, que nous manquons de ressources pour affronter le monde. Inversement, être d'humeur joviale, c'est se sentir disposé à s'amuser de différentes choses, et être d'humeur sereine, c'est sentir que les événements n'auront que peu de prise sur nous. À cet égard, les humeurs constituent probablement sur le plan biologique un mécanisme de régulation affective, à l'interface de nos désirs et de ce qui est susceptible de nous procurer du plaisir ou de la douleur, mécanisme dont le but serait de déterminer l'allocation de nos ressources énergétiques pour nos comportements (Grinde, 2012).

Il ne faudrait cependant pas exagérer ce premier contraste entre les humeurs et les émotions. De fait, les humeurs semblent également tournées vers le monde extérieur. Toutefois, là où les émotions portent sur un objet en particulier (ma tristesse porte sur la disparition de mon chat), les humeurs tendent à colorer d'une même teinte affective et phénoménale le monde qui nous entoure, et elles nous disposent à leur tour à éprouver certaines émotions. Bien plus, nos humeurs s'accompagnent en principe de certaines pensées caractéristiques à l'égard de leurs objets : l'anxiété nous dispose à envisager différentes formes de menace, l'agacement nous dispose à envisager certaines formes de ripostes, et ainsi de suite. Or il faut remarquer que c'est également vrai du bonheur : être heureux nous dispose typiquement à éprouver certaines émotions positives et à entretenir certaines pensées. Nous éprouverons plus facilement de l'amusement, de la joie ou de la tendresse en étant heureux, et nous serons ainsi plus disposés à répondre avec bienveillance à certaines situations.

En tenant compte de ces différentes considérations, plusieurs théories ont défendu l'idée que les humeurs possèdent un caractère adaptatif, dans la mesure où elles nous permettent de coordonner un ensemble de réponses affectives et comportementales déterminées en fonction de la nature de notre relation avec notre environnement. Larsen (2000), en particulier, a proposé une métaphore intéressante avec le modèle du « thermostat ». Selon ce modèle, nos humeurs servent à déclencher des changements organiques et cognitifs pour retrouver un état affectif positif standard, de même qu'un thermostat déclenche des changements pour ramener la température d'une pièce à la valeur souhaitée. Cette idée se marie particulièrement bien avec la thèse psychologique selon laquelle il existerait pour chaque individu un « point fixe de bonheur » (*happiness set point*) auquel il revient naturellement, c'est-à-dire un état de bonheur moyen qui constitue pour chacun une forme d'état affectif positif standard (Diener *et al.*, 2006). De fait, les études indiquent également que les individus sont généralement d'une humeur plutôt positive ou plaisante, tant qu'ils ne rencontrent aucun événement affectivement chargé et qu'ils se sentent ainsi ouverts à différentes sortes d'activités (Diener *et al.*, 2015).

Ces différentes caractéristiques rendent ainsi plausible l'idée que le bonheur pourrait être compris comme une humeur positive. Il est tout à fait envisageable que l'état que nous décrivons à travers le concept de bonheur ne désigne rien d'autre que l'ensemble des humeurs positives que nous sommes susceptibles d'éprouver lorsque les choses vont globalement bien pour nous : être d'humeur joviale, gaie, tranquille, enjouée, allègre, légère, etc. Toutes ces humeurs sont possiblement des instances de ce que nous appelons être heureux. Nous sommes heureux, en ce sens, lorsque nos humeurs sont suffisamment positives pour nous permettre d'échapper à une forme de neutralité affective, neutralité qui désignerait un point inférieur à la valeur attendue sur notre thermostat émotionnel. Cette manière de comprendre le bonheur est manifestement réductionniste, puisque le bonheur se réduirait en réalité aux différents types d'humeurs positives que nous pouvons éprouver. Il ne semble pas, en effet, que le bonheur coïncide avec un type de bonne humeur en particulier, puisque l'on peut être heureux en étant dans n'importe laquelle des humeurs positives que nous venons de mentionner⁹.

On peut toutefois être tenté de résister spontanément à l'idée que le bonheur ne soit rien d'autre qu'une humeur positive. Les humeurs sont en effet versatiles, dans la mesure où elles sont tributaires de nos émotions, et le bonheur semble parfois désigner quelque chose de plus global que notre humeur. Lorsque l'on demande à une personne si elle est heureuse « en ce moment », on ne souhaite pas seulement savoir comment elle va à l'instant même où nous lui posons la question, mais sur une période qui entoure ce moment et qui se compte plutôt en jours ou en semaines. Ainsi, la proposition alternative que défend Haybron (2008) revient à affirmer que le bonheur doit se caractériser de façon plus globale comme une propension à éprouver des humeurs et des émotions positives sur le long terme. En effet, les humeurs sont changeantes et Haybron soutient que lorsque nous nous demandons si une personne est heureuse, nous souhaitons savoir de façon plus générale si elle est disposée à ressentir des humeurs et des émotions positives (2008, p. 135).

De ce point de vue, le bonheur se définirait non plus comme un état ressenti, mais comme une disposition à ressentir certains états : « Être heureux [...] n'est pas nécessairement se *sentir* heureux. Le bonheur ainsi conçu n'est pas du tout une émotion ou une humeur particulière, mais consiste plutôt dans l'état ou la condition émotionnelle générale d'un sujet. » (Haybron, 2008, p. 66) En employant l'expression de « condition émotionnelle », Haybron cherche à exprimer l'idée que le bonheur aurait non seulement un aspect ressenti, mais aussi une dimension inconsciente renvoyant à des dispositions qui ne sont pas nécessairement ressenties. Le bonheur serait ainsi l'état émotionnel global d'une personne, au sens où être heureux consisterait à être dans une condition émotionnelle d'ensemble positive¹⁰. Cette condition émotionnelle se caractériserait selon Haybron à la fois par une forme d'« affirmation psychique » qui nous permet de répondre favorablement aux différentes circonstances de notre vie, mais aussi par une profondeur causale qui serait le corrélat de cette attitude affirmative. Être heureux, c'est être disposé causalement, par certaines propensions

émotionnelles, à pouvoir répondre positivement aux situations qui m'affectent. Ces propensions émotionnelles désigneraient des « états affectifs centraux », qui se distingueraient des états périphériques par leur persistance (être régulièrement de bonne humeur et non pas accidentellement), par un caractère diffus colorant la totalité de notre expérience, ou encore par une profondeur que n'ont pas tous les états plaisants, « [ce] qui distingue premièrement les états affectifs centraux [...] c'est qu'ils *disposent* les agents à faire l'expérience de certains affects plutôt que d'autres » (Haybron, 2008, p. 130). Le bonheur, selon Haybron, se caractériserait donc moins comme une humeur positive que comme un ensemble de dispositions à ressentir certaines humeurs et émotions positives de façon régulière.

Mais de nombreux points de cette approche sont discutables. La première difficulté vient du fait que Haybron ne précise pas vraiment ce que sont ces dispositions à ressentir des humeurs. Comme les humeurs sont déjà elles-mêmes des dispositions à ressentir des émotions, la théorie semble faire référence à une disposition affective d'arrière-plan plus générale que les humeurs. On pourrait alors imaginer que ces dispositions à éprouver des humeurs désignent le tempérament ou le caractère d'une personne, mais Haybron refuse explicitement cette assimilation. La condition émotionnelle d'un individu désigne quelque chose de « plus profond qu'une humeur ou une émotion, mais plus variable qu'un trait » (Haybron, 2010, p. 20). Mais il est difficile de voir à quel type de propriété psychologique cette caractérisation pourrait renvoyer.

De plus, y a-t-il vraiment un sens à concevoir le bonheur comme un épisode dont la durée serait supérieure à celle de nos humeurs ? Sur quelle durée faut-il alors se baser exactement pour mesurer un tel état : trois jours, une semaine, un mois, un trimestre ? Bien que cet usage du concept de bonheur renvoie probablement à une intuition communément partagée, il se peut aussi que cette intuition doive être considérée avec prudence. On peut être heureux plusieurs jours en raison d'expériences positives ou de l'absence d'accrocs majeurs dans notre quotidien, puis être dévasté par le décès d'un proche durant plusieurs semaines, et connaître à nouveau des jours heureux. Il est même possible d'être heureux toute une matinée, puis d'être profondément contrarié par des questions liées au travail ou à notre vie familiale, et retrouver la voie de la tranquillité quelques heures plus tard. Il n'y aurait pourtant aucun sens à dire que nous étions heureux dans les moments négatifs que nous avons traversés dans ces deux exemples. Ce que nous appelons être heureux, c'est aussi de ne pas être perturbé, agacé, envahi par le chagrin, l'insatisfaction, les tracas et les ennuis. C'est précisément ce ressenti que la thèse du bonheur comme bonne humeur s'efforce de saisir. De fait, c'est aux humeurs des individus que l'on se réfère pour évaluer leur état actuel, et il semble conceptuellement improbable qu'un individu soit d'humeur noire ou déprimée et qu'il prétende en même temps être heureux. D'un autre côté, il serait pour le moins étrange qu'une personne soit de bonne humeur mais qu'elle affirme en même temps être malheureuse.

Ainsi, le concept d'humeur permet de faire sens de l'intuition selon laquelle le bonheur désignerait un état plus ou moins durable. Mais cette même intuition semble conduire certains auteurs à imaginer que le bonheur serait fondamentalement irréductible à nos états affectifs, et notamment à nos humeurs. C'est en effet cette intuition que Haybron cherche à développer lorsqu'il défend l'idée que le bonheur consiste en une forme « d'affirmation psychique », qui serait plus ou moins composée de trois états : celui d'« harmonie » (*attunement*), celui d'« engagement » (*engagement*) et celui d'« approbation » (*endorsement*). Un individu heureux serait ainsi selon Haybron un individu confiant et ouvert au monde (harmonie), doté d'une certaine énergie dans le cadre des activités qu'il accomplit (engagement), et généralement joyeux (approbation). Mais cette caractérisation ressemble précisément à une idéalisation des personnes heureuses, plutôt qu'au portrait de personnes réellement heureuses. En réalité, le bonheur peut avoir différents profils qui n'impliquent pas nécessairement la panoplie complète de l'« affirmation psychique » : on peut être d'humeur enthousiaste, mais ne pas nécessairement être confiant ; d'humeur badine sans se sentir engagé ; d'humeur joyeuse sans être pour autant dans l'approbation. Les conditions de réalisation du bonheur, de ce point de vue, sont plurielles. Il existe différentes façons d'être heureux, qui dépendent de la variété de nos humeurs positives.

Le deuxième problème dans l'idée de propension est que le bonheur n'est plus caractérisé comme un état ressenti ou occurrent. Ce point est censé si l'on admet l'intuition de la durabilité. Mais il n'en demeure pas moins étrange de caractériser le bonheur comme un état qui ne serait pas nécessairement ressenti. S'il s'agit de caractériser l'existence d'un individu indépendamment de ce qu'il ressent, comme le suggère Haybron, alors on pourrait être tenté de penser que cette analyse renvoie en réalité au bien-être comme concept objectif et normatif, et qu'elle cherche avant tout à désigner ce qu'est une vie bonne et sous quelles conditions on peut l'atteindre. Inversement, le bonheur semble être davantage un concept subjectif : on se sent ou non heureux et il semblerait contradictoire qu'une personne puisse affirmer qu'elle est heureuse sans ressentir aucun état affectif, ou inversement que l'on puisse dire qu'une personne est objectivement heureuse alors qu'elle ne ressent aucun bonheur. Une autre option pour le partisan de la caractérisation du bonheur en termes de propensions pourrait consister à dire que le bonheur est ressenti lorsque nos propensions s'actualisent sous la forme d'humeurs et d'émotions. Mais cela revient alors à admettre que le bonheur n'est vraiment ressenti que sous la forme de ces états affectifs. Et comme nous avons déjà écarté l'approche émotionnelle du bonheur, alors la seule option plausible revient à affirmer que le bonheur est une humeur ressentie. On peut ressentir un bonheur plus ou moins intense, selon que notre humeur elle-même est particulièrement vive et qu'elle résulte de jugements ou d'émotions qui revêtent un caractère important ou « central » pour nous (pour paraphraser Haybron), mais c'est un point sur lequel nous reviendrons plus bas.

3. LES HUMEURS COMME ATTITUDES PROSPECTIVES

Si j'ai donné jusqu'ici un certain nombre de raisons permettant de penser que le bonheur consiste bien en une humeur positive, il reste néanmoins à proposer une théorie des humeurs qui serait susceptible de rendre compte de cette idée de façon convaincante. Dans cette section, je souhaiterais donc m'interroger sur la manière dont nous pouvons conceptualiser les humeurs pour proposer une théorie satisfaisante du lien qui unit les humeurs au bonheur. Or, il existe un fossé au sein des théories contemporaines entre deux types d'approches au sujet des humeurs. Je présenterai brièvement ces deux types de théories, avant d'en proposer une troisième, encore non explorée à ce jour, que je crois être plus satisfaisante pour rendre compte du rapport que les humeurs entretiennent avec le bonheur.

Selon certains philosophes, le caractère diffus des humeurs implique qu'elles seraient dépourvues d'intentionnalité. Ils proposent ainsi de comprendre les humeurs comme des états affectifs bruts qui coloreraient notre expérience avec certaines teintes phénoménales, mais qui seraient dépourvus d'objets (Roberts, 2003 ; Kind, 2013). Les humeurs ne feraient qu'exprimer nos états affectifs internes, sans être dirigées vers le monde extérieur. Plus généralement, elles ne seraient que le résultat de mécanismes causaux imperméables aux raisons : on est de bonne ou de mauvaise humeur « tout court » et non pas pour de bonnes ou de mauvaises raisons. Il ne semble pourtant pas si évident que les humeurs soient complètement étanches à l'égard des raisons. Certaines stratégies pour réguler nos humeurs sont, par exemple, liées aux raisons que nous avons ou non d'éprouver ces humeurs (Larsen, 2000). Par ailleurs, il fait partie de nos intuitions courantes que nous pouvons contrôler nos humeurs, tout comme nos émotions (Beedie, Terry et Lane, 2005 ; Cusimano et Goodwin, 2019). C'est aussi parce que nos humeurs sont sujettes à des contraintes de rationalité que certaines d'entre elles nous paraissent anormales, notamment lorsqu'elles sont déliées de toute référence à notre situation : il ne fait aucun sens d'être d'humeur déprimée si nous n'avons reçu aucune mauvaise nouvelle, d'être d'humeur colérique si nous n'avons rencontré aucun obstacle, ou d'être d'humeur rayonnante si les coups du sort s'acharnent sur nous.

Inversement, d'autres auteurs soutiennent que les humeurs sont des états intentionnels, au sens où elles possèderaient bien un objet : elles seraient dirigées vers le monde pris « comme un tout » (Solomon, 1993, p. 71). Cette seconde approche permet de proposer une explication intuitive du lien qui unit la phénoménologie des humeurs et leur intentionnalité : lorsque l'on est d'humeur noire, en effet, le monde entier semble offensant, et lorsque l'on est d'humeur badine, tout événement semble pouvoir être amusant. Une manière intéressante de comprendre l'intentionnalité des humeurs consiste alors à l'analyser en termes évaluatifs. Selon cette hypothèse, les humeurs sont des états évaluatifs dirigés vers le monde dans sa globalité. Tandis que les émotions seraient des évaluations de certains objets particuliers, les humeurs seraient des évaluations du monde en général. Tappolet (2018) soutient plus exactement que les humeurs seraient diri-

gées vers des « possibilités évaluatives ». Être d'humeur anxieuse reviendrait à percevoir la possibilité d'un danger. Dans la mesure où les humeurs visent des possibilités, il n'est pas nécessaire de supposer qu'elles portent sur un objet en particulier. On se donne ainsi les moyens de rendre compte à la fois de la généralité des humeurs et du fait qu'elles peuvent très bien ne pas être dirigées vers un objet. Enfin, les humeurs ont également des conditions de correction, puisqu'il semble correct d'être anxieux dans un environnement qui est effectivement chargé de menaces, et inversement.

En dépit des avantages que présente cette seconde approche, il faut reconnaître qu'il est toutefois difficile de concevoir les humeurs comme des représentations de valeurs ou de possibilités évaluatives. En effet, les humeurs semblent porter avant tout sur la diversité des objets qui nous entourent et dont elles sont susceptibles de s'emparer pour leur attribuer une coloration qui leur est propre. Autrement dit, les valeurs semblent moins être inhérentes au contenu représentationnel des humeurs, qu'à la manière dont nos humeurs elles-mêmes appréhendent ce contenu. À cet égard, une critique courante contre les théories perceptuelles des émotions semble également pouvoir être adressée aux théories perceptuelles des humeurs. Selon cette critique, si les humeurs étaient des perceptions évaluatives, alors un même contenu représentationnel devrait engendrer des représentations de valeurs ou de possibilités évaluatives identiques (Kind, 2013). Or ce n'est généralement pas le cas : deux individus peuvent entretenir des humeurs différentes vis-à-vis d'un même contenu représentationnel, et un même individu lui-même peut encore avoir des humeurs différentes vis-à-vis de ce même objet à différents moments. En ce sens, l'humeur semble bien plutôt consister en une manière d'évaluer un contenu représentationnel donné, qui n'est pas lui-même porteur de valeur en tant que tel (Dokic et Lemaire, 2013). De fait, si l'objet de mes humeurs est mon rapport général au monde, alors il est clair que cet objet peut être évalué de multiples façons à divers moments de mon histoire personnelle.

Ainsi, une manière d'analyser les humeurs, encore non explorée dans la littérature actuelle, pourrait consister à concevoir ces dernières comme des attitudes évaluatives spécifiques. L'idée d'analyser certains états affectifs comme des attitudes évaluatives a été proposée par Deonna et Teroni (2012) au sujet des émotions et elle repose sur une distinction entre l'attitude et le contenu caractéristiques d'un état mental. Le contenu d'un état mental peut avoir différents formats représentationnels : il peut être propositionnel ou non propositionnel par exemple (ce peut être une phrase tout comme la vision d'un objet ou l'audition d'un son). Et l'attitude désigne quant à elle la manière dont nous allons nous rapporter à ce contenu. On peut ainsi croire, supposer, désirer, ou espérer que « la France va battre les États-Unis au prochain mondial de basket ». Selon Deonna et Teroni, les émotions seraient des attitudes évaluatives en ce qu'elles consisteraient en des évaluations de certains contenus, qui se traduiraient notamment par des sensations corporelles de préparation à l'action (approcher ou fuir par exemple). Les sensations corporelles constitueraient ainsi la phénoménologie de l'émotion, et cette phénoménologie reflèterait l'intentionnalité évaluative de

l'émotion : une évaluation émotionnelle positive pourrait se caractériser par un rapprochement vis-à-vis de l'objet de l'émotion, et une évaluation émotionnelle négative par l'évitement de l'objet de l'émotion. L'évaluation ne serait donc pas dans le contenu, mais dans l'attitude spécifique qui se rapporte à ce contenu.

De la même manière, il est plausible de concevoir les humeurs comme des attitudes évaluatives, au sens où celles-ci ne semblent pas avoir un contenu représentationnel précis, mais consistent plutôt en des manières d'appréhender de façon plus générale non pas un contenu mental en particulier, mais les différents types de contenus que nous sommes susceptibles de rencontrer. En ce sens, les humeurs seraient des attitudes consistant à appréhender un ensemble d'événements différents, là où les émotions sont des évaluations spécifiques de certains objets. C'est à cette différence d'intentionnalité que l'on peut rattacher également la manière dont la phénoménologie des humeurs se distingue de celle des émotions. En effet, la phénoménologie des humeurs implique généralement moins de sensations corporelles que celle des émotions et elle est en ce sens plus évanescente que cette dernière. C'est essentiellement la composante hédonique des humeurs qui est ressentie, sous la forme d'un état plaisant ou déplaisant qui peut être plus ou moins prononcé. En revanche, cette phénoménologie comporte généralement une dimension cognitive prononcée, que ce soit par la manière dont elle mobilise notre focalisation attentionnelle, ou encore par le type de pensées qu'elle va impliquer (que ce soit sous la forme de ruminations ou de réflexions). Ainsi, là où les émotions se présentent comme une forme de préparation à l'action, les humeurs se présenteraient avant tout comme une forme de préparation mentale à appréhender certains événements, avec une dimension anticipatoire bien plus marquée (Price, 2006).

La question est alors de savoir ce en quoi consiste plus précisément le caractère évaluatif des humeurs, et en quoi il se distingue de celui des émotions. Comme nous l'avons déjà vu, les humeurs tendent à hériter leurs évaluations caractéristiques de nos émotions. Une émotion de colère intense pourra susciter une humeur noire, et une émotion de joie intense pourra inversement susciter une humeur exaltée. Les émotions, de leur côté, semblent être déclenchées par le fait que certains événements soient susceptibles de satisfaire ou de frustrer nos états motivationnels, tels que nos désirs ou nos attachements à certaines personnes, à certains objets, ou à des valeurs auxquelles nous tenons (Scherer, 2001). Je ressens de la fierté lorsque je parviens à accomplir un but que je me suis fixé, de la honte lorsque je ne suis pas à la hauteur de mes propres standards, de la tristesse lorsque je perds un bien auquel je tenais énormément. Nous éprouvons ainsi une grande diversité d'émotions au quotidien, et nos humeurs sont plus ou moins positives en fonction de la diversité de ces émotions et de leur intensité. On peut être de bonne humeur parce qu'une excellente nouvelle nous procure une joie plus intense que les petites contrariétés que nous avons rencontrées dans la journée ou dans la semaine, ou inversement de mauvaise humeur parce qu'une succession d'ennuis n'a pas été compensée par un ou plusieurs états positifs suffisamment intenses.

Une façon plausible d'analyser la relation entre les émotions et les humeurs qui en résultent pourrait ainsi consister à dire que les humeurs dépendent de la manière dont nous sommes émotionnellement touchés par ce qui nous importe. Autrement dit, les humeurs positives seraient des états de satisfaction ressentis causés par le fait que nos motivations sont « satisfaites » ou simplement préservées (comme lorsque mon sentiment amoureux est entretenu par une sortie au cinéma, une longue discussion intime, ou une surprise de la part de ma compagne). Mais si les humeurs sont causées par nos émotions, elles semblent toutefois davantage tournées vers le monde extérieur, en prenant la forme d'attitudes « prospectives » : elles évaluent la possibilité plus générale que certains événements puissent affecter positivement ou négativement nos motivations. Ce caractère prospectif permet justement d'expliquer pourquoi les humeurs ont cette dimension modale qui est bien mise en avant par la théorie de Tappolet. Nos humeurs sont, pour ainsi dire, aux aguets des différentes possibilités de satisfaction ou de contrariété que pourraient rencontrer nos motivations. La prospection, en effet, se caractérise comme une « représentation de futurs possibles », pour reprendre l'expression de Seligman et ses collègues (2013)¹¹, représentation dont la fonction est de générer et d'évaluer sous forme de conditionnels différents états possibles. Comme le disent encore Seligman et ses collègues, « l'organisme prospecteur doit construire un *paysage évaluatif* des actions et de leurs conséquences possibles » (2013, p. 120). En ce sens, la prospection témoignerait selon eux d'une forme d'intelligence implicite dans laquelle nos états affectifs nous permettraient précisément d'évaluer les simulations de différents états futurs, pour nous préparer et nous adapter aux changements possibles auxquels nous pourrions être confrontés. Même les humeurs comme la mélancolie, de ce point de vue, qui semblent davantage focalisées sur le passé, pourraient consister en réalité à envisager différentes réécritures possibles de notre propre histoire¹².

On pourrait dès lors suggérer, à la suite de Nesse (1991), que les humeurs évaluent la « propiciosité » des situations auxquelles nous faisons face. Selon cette approche, les humeurs positives se caractérisent par le fait qu'elles vont mobiliser des ressources cognitives, physiologiques et comportementales de notre organisme afin d'investir une certaine somme d'efforts face à une situation qui est évaluée comme étant propice à un tel investissement. Inversement, les humeurs négatives se caractérisent par un retrait de ce genre d'investissement face à une situation qui est évaluée comme non propice. Autrement dit, les humeurs comme attitudes prospectives se caractériseraient par des évaluations qui mobilisent l'allocation de nos ressources organiques et cognitives en fonction de la probabilité que cet investissement soit « rémunérateur », et qu'il puisse ainsi nous permettre de satisfaire nos motivations. De telles évaluations impliquent ainsi la probabilité que nous rencontrions des dangers dans le cas d'une humeur anxieuse, que nous puissions nous réjouir dans le cas d'une humeur joviale, ou que nous rencontrions des obstructions dans le cas d'une humeur colérique. Plus nous évaluons les situations comme étant propices à nos efforts, plus notre humeur sera positive et nous disposera à mobiliser davantage de ressources cognitives, physiologiques, comportementales, et inversement

lorsque la propiciosité semble négative. On est « d'humeur à » travailler précisément parce qu'on évalue la situation comme étant propice à ce que nos efforts dans le travail soient payants. C'est en ce sens que nos humeurs semblent influencer notre motivation au quotidien et nos différents états émotionnels.

Notons d'emblée qu'une telle approche permet de mettre l'accent sur le caractère adaptatif des humeurs, ainsi que sur leur caractère diffus et envahissant, dans la mesure où elle insiste sur les relations que les humeurs entretiennent avec nos autres états mentaux, et notamment nos émotions. Cette analyse permet également d'insister sur la nature des attitudes évaluatives qui caractérisent les humeurs, en les comprenant comme des formes de mobilisation ou de démobilisation ressenties, à la fois du point de vue corporel et du point de vue cognitif. Enfin, cette approche permet de comprendre en quoi les humeurs comme attitudes peuvent être correctes ou incorrectes. En effet, si les humeurs portent sur la propiciosité des situations auxquelles nous faisons face, alors il semble incorrect d'être de bonne humeur dans des situations qui ne sont pas propices à nos motivations, et inversement. Mais la propiciosité, ici, ne renvoie pas uniquement à des « possibilités évaluatives » objectives, comme c'était le cas dans l'approche de Tappolet, puisque le caractère propice d'une situation dépend aussi de nos préférences subjectives, ainsi que de l'énergie dont nous disposons pour saisir certaines occasions. Il semble ainsi correct d'être d'humeur déprimée après la mort d'un proche, même s'il n'existe aucune possibilité sérieuse que nous ne perdions quelqu'un d'autre actuellement, tout simplement parce nous n'avons plus la force d'affronter le monde sans cette personne.

Résumons l'approche attitudinale des humeurs, afin de mettre au jour les liens qui unissent cette approche avec la thèse plus générale selon laquelle le bonheur est une humeur positive. Selon la thèse attitudinale défendue dans cette section, les humeurs sont des attitudes évaluatives prospectives : elles évaluent la propiciosité des situations que nous rencontrons, et ces évaluations se traduisent par une phénoménologie affective et cognitive. Cette appréciation de la propiciosité dépend le plus souvent des émotions que nous éprouvons et qui engendrent à leur tour nos humeurs, mais elle peut dépendre également de nos autres états mentaux (un raisonnement sur notre situation actuelle, une réflexion sur notre avenir, etc.). Il devient intuitif, dès lors, de considérer que le bonheur n'est peut-être rien d'autre qu'une humeur positive, c'est-à-dire une évaluation positive de la propiciosité des situations que nous rencontrons.

Cette approche permet de rendre compte du fait que le bonheur désigne une attitude plus générale qu'une simple émotion : les humeurs dérivent en effet, comme nous l'avons vu, de la diversité de nos expériences émotionnelles et du rapport de celles-ci avec nos motivations. En ce sens, nos humeurs constituent une espèce d'état affectif intermédiaire, dans lequel se cristallisent nos autres états affectifs liés à la satisfaction ou à la frustration de nos motivations. Ce point apparaissait déjà lorsque nous avons souligné, dans la section précédente, le rôle régulateur des humeurs, et c'est un aspect dont il devient facile de rendre compte en insistant sur la manière prospective dont nos humeurs évaluent la

propiciété des situations. En effet, les humeurs régulent nos investissements motivationnels en fonction de la manière dont elles évaluent la probabilité qu'un désir ou une préférence puisse être satisfait, notamment en regard de nos expériences passées et des émotions que celles-ci ont engendrées.

Ce même point permet de rendre compte de la diversité des aspects que Haybron mentionnait au sujet du bonheur. En effet, nous pouvons nous sentir engagés avec le monde lorsque notre humeur est liée aux émotions positives que nous procure une activité, ou bien dans un état d'approbation lorsque nous éprouvons une forme de joie dans l'accomplissement d'un désir qui trouve enfin satisfaction, ou encore nous sentir en harmonie avec le monde lorsque nous sommes d'humeur optimiste, parce que nous avons bon espoir que nos désirs seront satisfaits. Mais il est possible encore d'imaginer bien d'autres humeurs positives constitutives des différents épisodes de bonheur que nous sommes susceptibles de traverser, notamment en lien avec nos relations interhumaines, ou avec nos expériences esthétiques.

Il devient aussi possible, à partir de là, de saisir de façon un peu moins énigmatique une idée que nous avons croisée dans les approches émotionnelles du bonheur, à savoir le fait que le bonheur est une attitude tournée vers notre bonne fortune, dans la mesure où les humeurs consistent en des appréhensions générales du caractère propice ou non propice de notre situation. Nous sommes de bonne humeur parce que les choses vont globalement bien pour nous, en regard de nos émotions et de ce à quoi nous tenons, et cela, il est vrai, a probablement grandement à voir avec la chance. De façon plus prosaïque, l'idée que les humeurs seraient des évaluations de la propiciété de notre situation revient littéralement à dire qu'elles évaluent les chances que nous avons de pouvoir satisfaire ou non nos motivations.

4. HUMEURS ET BIEN-ÊTRE

Dans cette dernière section, je souhaite examiner les relations qui unissent la théorie du bonheur comme humeur positive avec le concept de bien-être. Selon les théories du bonheur au sujet du bien-être, seul le bonheur est non instrumentalement bon pour un sujet (Fletcher, 2016). Une manière intuitive de développer cette thèse en adoptant l'approche du bonheur comme bonne humeur consiste à soutenir que les autres ingrédients que l'on conçoit parfois comme des constituants du bien-être (tels que le plaisir, la satisfaction du désir, ou encore certains biens « objectifs » tels que l'amitié ou la santé) ne sont bons que parce qu'ils permettent d'engendrer des humeurs positives.

Deux raisons militent en faveur de cette thèse. La première est que l'on peut éprouver autant de plaisirs que l'on souhaite, satisfaire autant de désirs que possible et jouir de nombreux biens supposément objectifs, il serait difficilement concevable d'admettre que nous allons bien si tout cela ne nous conduisait pas à être de bonne humeur. Les plaisirs et les désirs satisfaits contribuent à notre bien-être dans la mesure où ils contribuent à engendrer des humeurs positives.

Inversement, si les humeurs positives sont bonnes pour nous, ce n'est pas seulement parce qu'elles sont plaisantes, mais parce que ce caractère plaisant se traduit sous la forme d'attitudes caractéristiques, qui nous rendent plus ouverts et accueillants à l'égard de notre relation avec le monde. Ce ne sont pas les plaisirs ou la satisfaction de nos désirs que nous recherchons ultimement, mais les bonnes humeurs comme attitudes qui en découlent.

La deuxième raison est un corollaire de la première : soutenir que d'autres biens sont non instrumentalement bons même s'ils ne nous conduisent pas à être de bonne humeur semble conduire à une théorie du bien-être probablement aliénante. Admettons par hypothèse, en effet, que la santé, l'autonomie, ou l'amitié, par exemple, sont des biens dotés d'une valeur prudentielle indépendamment de notre attitude subjective à leur égard. Il serait étrange, comme on le souligne régulièrement depuis Railton (2003), de faire dépendre ainsi notre bien-être de biens extérieurs qui n'ont peut-être pas toujours une influence sur ce dernier. Il est même concevable que ces biens ne soient pas nécessairement bons pour tous les sujets et que certains d'entre eux laissent certains d'entre nous indifférents. Il serait problématique qu'une théorie du bien-être ne soit pas sensible aux particularités et aux idiosyncrasies affectives de chacun. Dans la mesure où nos humeurs sont tributaires de nos réactions émotionnelles, il est peu probable qu'un bien incompatible avec nos préférences et nos motivations subjectives puisse nous rendre heureux et être ainsi ultimement bon pour nous.

La théorie des humeurs au sujet du bien-être présente donc plusieurs avantages qui lui donnent de la crédibilité à première vue. Elle est en mesure de rendre compte des raisons pour lesquelles certains biens subjectifs (tels que les plaisirs ou la satisfaction du désir) sont bons pour un individu. Mais elle est aussi capable de rendre compte des raisons pour lesquelles les biens prétendument objectifs sont parfois bons pour nous. Ils le sont en effet lorsqu'ils sont compatibles avec notre tempérament, ou lorsque la privation de ces biens peut avoir des effets délétères sur nos humeurs : le sommeil ou la santé sont ainsi des biens objectifs pour cette raison, dans la mesure où ils ont une valeur instrumentale dans la recherche du bien-être. Enfin, cette théorie du bien-être permet de rendre compte de l'intuition courante selon laquelle nous allons bien lorsque notre situation est globalement bonne. Il ne suffit pas, en effet, d'être ponctuellement de bonne humeur pour aller bien. Un haut niveau de bien-être suppose la continuité et la succession de nos humeurs positives. En ce sens, les bonnes humeurs sont des constituants de notre bien-être plus que des causes. Notre niveau de bien-être n'est pas déterminé seulement par notre humeur actuelle, mais par l'ensemble des humeurs positives que nous éprouvons en moyenne.

Ce dernier point est important, car il permet de prendre en charge différentes objections que l'on pourrait adresser à la théorie du bonheur comme bonne humeur. On pourrait par exemple objecter que l'on peut être parfois de bonne humeur tout en étant globalement malheureux. Dans les termes de la théorie défendue ici, cela revient en fait simplement à dire que l'on peut éprouver des moments de bonheur, dans une vie dont le niveau de bien-être est globalement

faible, voire négatif. Inversement, on pourrait objecter que l'on peut rester globalement heureux même quand des petits tracasseries nous rendent de mauvaise humeur. Mais là encore, cela revient en réalité à dire que l'on peut jouir d'un haut niveau de bien-être parce que l'on est fréquemment de bonne humeur, tout en traversant de brefs épisodes malheureux. Autrement dit, la thèse défendue ici nous invite à considérer toutes les situations où nous parlons de bonheur ou de malheur de façon « globale » comme des métaphores désignant en réalité le bien-être.

La question qui reste en suspens est alors de savoir si cette théorie est également susceptible de nous fournir une perspective critique sur le bien-être des individus. Après tout, les théories du bien-être sont censées être normatives, là où les théories du bonheur sont simplement descriptives (Haybron, 2008), et l'on attend précisément d'une théorie du bien-être qu'elle soit en mesure de nous dire dans quel cas certaines choses peuvent ne pas être bonnes pour un individu, quelle que soit par ailleurs l'attitude de celui-ci à l'égard de ces choses (Lemaire, 2016). Or, une objection évidente pourrait insister sur le fait que la théorie des humeurs au sujet du bonheur semble laisser croire qu'il suffit d'être de bonne humeur pour aller bien. Et manifestement, la plupart des gens sont souvent de bonne humeur (Diener *et al.*, 2015). Dans ce cas, la théorie perdrait au mieux toute force normative et se contenterait de constater ce qui n'est plus qu'un fait banal : la plupart des gens sont heureux et vont bien. Les choses sont pourtant loin d'être aussi simples, comme en témoignent l'augmentation importante des dépressions sévères (Wakefield et Demazeux, 2015), les troubles du sommeil de plus en plus fréquents (Walker, 2017) ainsi que les nombreuses pathologies qui accompagnent le monde du travail. Le fait que nous sommes en moyenne plutôt heureux ne devrait donc pas nous conduire à mettre de côté les nombreux cas dans lesquels nous cessons de l'être, souvent sous la pression de conditions extérieures difficiles qui viennent à leur tour affecter nos humeurs.

Une autre implication contre-intuitive de la théorie des humeurs au sujet du bien-être est la suivante. Puisque cette théorie affirme qu'il suffit d'être de bonne humeur pour aller bien, alors il suffirait selon cette théorie que chacun d'entre nous soit soumis à une consommation de drogue optimale pour avoir un niveau de bien-être suffisant, quelles que soient par ailleurs nos conditions de travail ou peu importe notre état général de santé. Cette conséquence serait pour le moins contre-intuitive et rendrait cette théorie beaucoup moins attrayante. Mais si les bonnes humeurs sont censées contribuer à notre bien-être, c'est précisément en vertu de leur nature évaluative. Or, une drogue qui nous conduirait à être toujours de bonne humeur reviendrait à perturber causalement le mécanisme évaluatif des humeurs et nous conduirait ainsi à des évaluations détachées de notre rapport réel au monde.

En ce sens, notre bien-être dépend d'un bonheur qui est lui-même approprié, c'est-à-dire d'humeurs positives qui évaluent correctement la propiété des situations que nous rencontrons. S'il n'est pas nécessaire d'éprouver des humeurs correctes pour être heureux (dans la mesure où le bonheur est un concept simplement descriptif), il semble en revanche crucial que notre bien-

être repose sur des évaluations correctes, sans quoi il serait effectivement possible de considérer n'importe quelle situation de bonheur comme dotée d'une valeur prudentielle. Mais il n'est guère concevable qu'un bonheur illusoire soit favorable à notre bien-être, qu'il soit lié à l'absorption de drogues ou à d'autres circonstances.

Il reste alors à définir plus précisément ce que désigneraient les conditions de correction de nos humeurs. Faute de place, je ne peux qu'esquisser ici ce qui me semble être une réponse plausible. Deux conditions au moins semblent importantes pour spécifier la correction des humeurs. La première est que nos humeurs, comme nous l'avons vu, sont connectées à nos autres états affectifs et motivationnels, tels que nos désirs, nos préférences ou nos sentiments. Une situation n'est propice, en ce sens, que si elle permet d'une façon ou d'une autre de satisfaire nos états motivationnels. L'évaluation de la propicuosité est donc indexée sur l'importance subjective que les situations ont pour nous, en regard de nos motivations. Il semble par exemple incorrect d'être de mauvaise humeur au sujet d'une situation objectivement défavorable, mais qui ne contrarierait aucune de nos motivations (être de mauvaise humeur à cause de la chute de la bourse, alors que cette situation n'est aucunement liée à nos préférences).

Une deuxième condition importante est liée à la centralité de nos états motivationnels, une caractéristique qui était également au cœur de l'analyse du bonheur proposée par Haybron, comme nous l'avons vu dans la section 2. L'importance subjective des situations, en effet, semble être liée au caractère plus ou moins central des motivations auxquelles elles renvoient. La victoire en championnat des Golden State Warriors est subjectivement moins importante pour moi que la santé de ma conjointe. Il serait donc incorrect d'être davantage affecté par la première que par la seconde. Un état motivationnel est dit « central », de ce point de vue, en ce qu'il est durable et plus étroitement connecté avec nos autres états mentaux, par contraste avec les états plus « périphériques » qui sont au contraire transitoires et moins connectés avec nos autres états mentaux (Arpaly et Schroeder, 1999). La victoire des Warriors, de ce point de vue, ne renvoie qu'à l'une de mes vagues préférences sportives, tandis que la santé de ma conjointe renvoie à mes intérêts et mes désirs les plus constants.

Cette caractérisation de la correction s'efforce de trouver un juste milieu entre les théories objectivistes, qui font dépendre notre bien-être de biens extérieurs, et les théories subjectivistes, qui font dépendre notre bien-être principalement de nos états internes, tels que nos plaisirs, nos désirs, ou notre bonheur. Une humeur est bonne pour nous tant qu'elle possède une valence positive et qu'elle évalue correctement une situation en regard de nos motivations. Cette théorie comporte une dimension subjectivement ressentie, mais elle implique aussi des conditions de corrections qui, même si elles dépendent de nos motivations, sont aussi indépendantes de nous (dans la mesure où il ne dépend pas de nous qu'un état du monde corresponde effectivement ou non à nos préférences). Ainsi, le fait d'éprouver des humeurs positives parce que l'on a consommé une certaine drogue ne suffit pas à rendre notre vie bonne. Il faut encore que le monde corres-

ponde effectivement à certaines de nos motivations et que nous évaluions correctement cette correspondance.

CONCLUSION

Les théories qui soutiennent que le bonheur est une émotion insistent sur la dimension ressentie du bonheur, mais ne donnent généralement pas de raison suffisante de penser que cette émotion serait réellement différente des émotions de joie ou de contentement et tendent à mettre de côté la dimension plus ou moins durable du bonheur. Inversement, les théories qui assimilent le bonheur à une forme de « condition émotionnelle » impliquant des propensions à être de bonne humeur tendent à mettre de côté la dimension ressentie du bonheur. Dans cet article, je me suis efforcé de montrer que l'on pouvait concevoir le bonheur plus simplement comme une forme d'humeur positive, qui est ressentie et dans laquelle viennent se généraliser nos évaluations émotionnelles et non émotionnelles.

J'ai proposé de comprendre les humeurs comme des évaluations de ce que Nesse nomme la « propiciosité » des situations. Nos humeurs évaluent la possibilité que ces situations nous permettent de satisfaire ou non nos états motivationnels. En ce sens, ce sont des attitudes prospectives, marquées par une dimension anticipatoire, qui nous permettent de déterminer s'il vaut la peine ou non d'investir une certaine somme d'efforts dans la poursuite d'un but donné. Nos humeurs, en effet, tiennent compte de nos différentes expériences émotionnelles, mais aussi de nos jugements, et elles constituent en ce sens un type de mécanisme affectif qui, plus général que nos émotions, nous permet d'adapter notre comportement aux changements possibles que nous sommes susceptibles de rencontrer. Être heureux, de ce point de vue, c'est évaluer les situations auxquelles nous faisons face comme étant propices à la satisfaction de nos motivations. On peut ainsi être heureux de différentes façons, en fonction de la diversité de nos humeurs positives, telles que l'enthousiasme, la joie, la sérénité, etc., qui sont toutes marquées par une phénoménologie différente (de l'entrain, de la satisfaction, de l'apaisement, etc.). Et on peut être plus ou moins heureux selon que le monde nous paraît plus ou moins propice.

À partir de cette analyse, j'ai proposé de comprendre le bien-être comme le fait d'éprouver des humeurs positives qui évaluent correctement la propiciosité des situations. Notre niveau de bien-être, en d'autres termes, est constitué par l'ensemble des humeurs positives correctes que nous éprouvons.

NOTES

- ¹ Goldman (2017) ; Mulligan (2016).
- ² Haybron (2008) ; Kauppinen (2013).
- ³ Voir toutefois Kahneman et Krueger (2006).
- ⁴ Il semble en revanche concevable qu'une personne puisse être de bonne humeur ponctuellement tout en étant globalement malheureuse. Je tâcherai de montrer dans la dernière section comment la théorie proposée ici peut rendre compte de ce genre de phénomène.
- ⁵ Dans ce qui suit, j'emploierai de façon indifférenciée les concepts d'« état motivationnel » et de « motivation ».
- ⁶ Si Goldman soutient qu'aucune propriété des émotions ne serait vraiment nécessaire ou suffisante pour les caractériser (2017, p. 7), il soutient en même temps que le jugement joue un rôle central (2017, p. 10).
- ⁷ Mulligan n'emploie pas l'expression d'« objet formel » dans ce texte, mais celle « d'objet impropre ». Il semble toutefois désigner par cette expression la même chose que ce que d'autres nomment « objet formel ». Par simplicité, je préfère donc conserver ce vocable.
- ⁸ Il ne s'agit pas ici de dire que l'étymologie est conceptuellement déterminante, mais de constater le simple fait que bonheur et chance sont, effectivement, couramment associés.
- ⁹ Comme l'a souligné un relecteur anonyme, cette thèse n'est toutefois pas nécessairement réductionniste *stricto sensu* si l'on admet que certains facteurs (par exemple des facteurs individuels liés à notre histoire, notre constitution, etc.) peuvent déterminer le seuil à partir duquel une humeur est suffisamment positive pour que nous soyons heureux.
- ¹⁰ Kauppinen (2013) adopte également une théorie similaire.
- ¹¹ Je remercie Dan Haybron d'avoir attiré mon attention sur ce texte.
- ¹² Je remercie Christine Tappolet d'avoir attiré mon attention sur ce cas problématique.

BIBLIOGRAPHIE

Arpaly, Nomy et Timothy Schroeder, « Praise, Blame, and the Whole Self », *Philosophical studies*, vol. 93, no. 2, 1999, p. 161-188.

Beedie, Christopher J., Peter C. Terry et Andrew M. Lane, « Distinctions Between Emotion and Mood », *Cognition and Emotion*, vol. 19, no. 6, 2005, p. 847-878.

Cusimano, Corey et Geoffrey P. Goodwin, « Lay Beliefs about the Controllability of Everyday Mental States », *Journal of Experimental Psychology : General*, vol. 148, no. 10, 2019, p. 1701-1732.

DeLancey, Crag, « Basic Moods », *Philosophical Psychology*, vol. 19, no. 4, 2006, p. 527-538.

Deonna, Julien A. et Fabrice Teroni, *The Emotions: A Philosophical Introduction*, London, Routledge, 2012.

Diener, Ed, Richard E. Lucas et Christie Napa Scollon, « Beyond the Hedonic Treadmill: Revising the Adaptation Theory of Well-Being », *American Psychologist*, vol. 61, no. 4, 2006, p. 305-314.

Diener, Ed, Satoshi Kanazawa, Eunkook M. Suh et Shigehiro Oishi, « Why People Are in a Generally Good Mood », *Personality and Social Psychology Review*, vol. 19, no. 3, 2015, p. 235-256.

Dokic, Jerome et Stéphane Lemaire, « Are Emotions Perceptions of Value? », *Canadian Journal of Philosophy*, vol. 43, no. 2, 2013, p. 227-247.

Goldman, Alan H., « Happiness is an Emotion », *The Journal of Ethics*, vol. 21, no. 1, 2017, p. 1-16.

Grinde, Bjørn, *The Biology of Happiness*, Dordrecht, Springer, 2012.

Haybron, Daniel M., *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, Oxford, Oxford University Press, 2008.

———, « Mood Propensity as a Constituent of Happiness: A Rejoinder to Hill », *Journal of Happiness Studies*, vol. 11, no. 1, 2010, p. 19-31.

Kahneman, Daniel et Alan B. Krueger, « Developments in the Measurement of Subjective Well-Being », *Journal of Economic Perspectives*, vol. 20, no. 1, 2006, p. 3-24.

Kauppinen, Antti, « Meaning and Happiness », *Philosophical Topics*, vol. 41, no. 1, 2013, p. 161-185.

Kind, Amy, « The Case Against Representationalism about Moods », in Uriah Kriegel (dir.), *Current Controversies in the Philosophy of Mind*, London, Routledge, 2013, p. 113-134.

Larsen, Randy J., « Toward a science of mood regulation », *Psychological Inquiry*, vol. 11, no. 3, 2000, p. 129-141.

Lemaire, Stéphane, « A Stringent but Critical Actualist Subjectivism about Well-Being », *Les ateliers de l'éthique/The Ethics Forum*, vol. 11, no. 2-3, 2016, p. 133-150.

Mulligan, Kevin, « Happiness, Luck and Satisfaction », *Argumenta*, vol. 1, no. 2, 2016, p. 133-145.

Nesse, Randolph M., « What Is Mood For? », *Psychology*, vol. 2, no. 9.2, 1991.

Price, Carolyn, « Affect without Object: Moods and Objectless Emotions », *European Journal of Analytical Philosophy*, vol. 2, no. 1, 2006, p. 49-68.

Railton, Peter, « Facts and Values », in Peter Railton, *Facts, Values and Norms: Essays toward a Morality of Consequence*, New York, Cambridge University Press, 2003, p. 43-68.

Roberts, Robert C., *Emotions: An Essay in Aid of Moral Psychology*, Cambridge, Cambridge University Press, 2003.

Scherer, Klaus R., « Appraisal Considered as a Process of Multi-Level Sequential Checking », in Klaus R. Scherer, Angela Schorr, et Tom Johnstone (dir.), *Appraisal Processes in Emotion: Theory, Methods, Research*, New York, Oxford University Press, 2001, p. 92-120.

Seligman, Martin, Peter Railton, Roy F. Baumeister et Chandra Sripada, « Navigating Into the Future or Driven by the Past », *Perspectives on Psychological Science*, vol. 8, no. 2, 2013, p. 119-141.

Siemer, Matthias, « Mood Experience: Implications of a Dispositional Theory of Moods », *Emotion Review*, vol. 1, no. 3, 2009, p. 256-263.

Solomon, Robert, *The Passions, Emotions and the Meaning of Life*, Indianapolis, Hackett Publishing Company, 1993.

Sumner, L.W., *Welfare, Happiness and Ethics*, New York, Oxford University Press, 1996.

Tappolet, Christine, *Emotions, Values and Agency*, New York, Oxford University Press, 2016.

———, « The Metaphysics of Moods », in Hichem Naar et Fabrice Teroni (dir.), *The Ontology of Emotions*, Cambridge, Cambridge University Press, 2018, p. 169-186.

Teroni, Fabrice, « Emotions and Formal Objects », *Dialectica*, vol. 61, no. 3, 2007, p. 395-415.

Wakefield, Jerome C. et Steeves Demazeux (dir.), *Sadness or Depression? International Perspectives on the Depression Epidemic and its Meaning*, Dordrecht, Springer, 2016.

Walker, Matthew, *Why We Sleep: The New Science of Sleep and Dreams*, Penguin, London, 2017.

THE HEDONIST'S EMOTIONS

JULIEN DEONNA

DEPARTMENT OF PHILOSOPHY, UNIVERSITY OF GENEVA

FABRICE TERONI

DEPARTMENT OF PHILOSOPHY, UNIVERSITY OF GENEVA

ABSTRACT:

This paper explores the compelling hedonistic intuition that emotions affect happiness because they are *states of pleasure and displeasure*. The discussion focusses on two constraints on a plausible version of hedonism and explains which accounts of the emotions satisfy these constraints. Section 1 revolves around the *nonalienation constraint*: the constituents of a subject's happiness must engage him or her. We argue that the intuition that emotions have prudential value presupposes that emotions are forms of engagement, a condition that only some accounts of the emotions satisfy. Section 2 centres around the *unity constraint*: if we acknowledge a great variety of (dis)pleasures, we still need to understand what makes all of them (dis)pleasures. Conceiving of the emotions as forms of engagement, we contend, allows us to resolve the difficulties concerning the variety and unity of (dis)pleasures that weigh on traditional hedonism. In section 3, we defend the form of affective hedonism that has emerged. We argue that the approach can be extended from emotions to other affective states and that the central role we give to action tendencies in our conception of affectivity does not call into question the idea that emotions contribute to happiness because of their hedonic value.

RÉSUMÉ :

Cet article explore l'intuition hédoniste convaincante selon laquelle les émotions affectent le bonheur parce *qu'elles sont des états de plaisir et de déplaisir*. La discussion s'intéresse à deux contraintes sur une version plausible de l'hédonisme et explique quels récits des émotions satisfont ces contraintes. La section 1 s'articule autour de la *contrainte de aliénation* : les constituants du bonheur d'un sujet doivent l'engager. Nous soutenons que l'intuition selon laquelle les émotions ont une valeur prudentielle présuppose que les émotions sont des formes d'engagement, une condition que seules certaines conceptions des émotions satisfont. La section 2 est centrée sur la *contrainte de l'unité*: si nous reconnaissons une grande variété de (dé)plaisirs, nous devons encore comprendre ce qui en fait des (dé)plaisirs. Concevoir les émotions comme des formes d'engagement, soutenons-nous, permet de résoudre les difficultés concernant la variété et l'unité des (dé)plaisirs qui pèsent sur l'hédonisme traditionnel. Dans la section 3, nous défendons la forme d'hédonisme affectif qui se dégage de ce qui précède. Nous soutenons que cette approche peut être étendue des émotions à d'autres états affectifs et que le rôle central que nous donnons aux tendances à l'action dans notre conception de l'affectivité ne remet pas en question l'idée selon laquelle les émotions contribuent au bonheur en raison de leur valeur hédonique.

INTRODUCTION

What is happiness made of? A fundamental intuition is that emotions and, more generally, affective states, contribute, for better or for worse, to the value of our lives. There are surely several reasons why emotions have value and, more specifically, why they have what is usually called prudential value—that is, the kind of value that allows them to contribute positively or negatively to the happiness of those who experience them. For example, emotions often inform us about the things that matter to us, prepare us to negotiate challenges that we face, and are central aspects of the relationships of love and friendship, but also of hate and rejection, at the heart of our existences (Deonna and Teroni, 2013; Haybron, 2008; Rossi and Tappolet, 2015). In what follows, we shall leave aside many contributions of emotions to happiness so as to focus on the traditional idea that emotions affect happiness because they are *states of pleasure and displeasure*. To this extent, the intuition that constitutes our starting point is hedonistic.¹ On the basis of this intuition, we wish to address two questions that have preoccupied hedonists and their critics.

First, what constraints must emotions satisfy in order to contribute hedonically to happiness? In the first part of our discussion, we start from the idea that the constituents of a subject's happiness must somehow engage him or her: they cannot leave the subject totally cold or be imposed on him or her from the outside. Let us call this the *nonalienation constraint* (Fletcher, 2013; Railton, 1986). On this basis, we shall try to show that the defence of the intuition that emotions have prudential value presupposes that they are forms of engagement, a condition that only some approaches to the emotions satisfy.

The second question we wish to address, which follows from the idea that different emotions are distinct types of engagement, is that of the unity and variety of states of (dis)pleasure (Clark, 2005; Feldman, 2004). If we acknowledge the existence of a great variety of (dis)pleasures, we still need to understand what makes all of them (dis)pleasures—otherwise the attractiveness of hedonism would diminish. To what extent are (dis)pleasures unified? Let us call this the *unity constraint*. In the second part of our discussion, we shall argue that conceiving of the emotions as forms of engagement allows us to resolve in an attractive way the difficulties concerning the variety and unity of (dis)pleasures that weigh on traditional hedonism.

In the last part, we shall defend the form of affective hedonism that will have emerged. In particular, we shall argue that the approach we advocate can be extended from emotions to other affective states and that the central role we give to action tendencies in our conception of affectivity does not call into question the idea that emotions contribute to happiness because of their hedonic value.

§1 A NONALIENATING CONCEPTION OF THE EMOTIONS

If something contributes to subjects' happiness, it cannot leave them totally cold and fail to engage them. As Peter Railton aptly puts it, "it would be an intolerably alienated conception of someone's good to imagine that it could fail in any way to engage him" (1986/2003, p. 47; see also Fletcher, 2013). In the hedonist framework we presuppose here, letting chocolate melt in one's mouth, reading a novel, listening to an opera, riding a bicycle up a mountain pass, and so on, contribute to happiness only to the extent that one takes pleasure or displeasure in such activities.

The same observation seems to apply to any bodily sensation: the sensation of chocolate melting in your mouth contributes positively to your happiness only if you take pleasure in it.² Consider the ascetic, a figure who is often said to have an unfavourable attitude toward pleasures. To the extent that the relevant sensations do not please the ascetic, as opposed to what is the case for most of us, these sensations do not contribute positively to the ascetic's happiness. The challenge for hedonism is to preserve the intuition that (dis)pleasures contribute to happiness, which requires the development of a conception of taking (dis)pleasure that engages the subject. This is the nonalienation constraint on the constituents of happiness.³

If we accept our initial intuition that emotions are constituents of happiness, this constraint must be met: emotions must constitute forms of commitment on the subject's part. Which approaches to the emotions make them relevant forms of commitment? We shall first argue that two classical theories of the emotions do not fit the bill: the Jamesian theory and the theory that conceives emotions as perceptions of values. On this basis, we shall argue that the nonalienation constraint supports theories that emphasize the close links between emotions and actions.

Let us first consider the Jamesian theory (James, 1890; see also Prinz, 2004). According to this theory, emotions are configurations of bodily sensations. These sensations are underpinned by changes in a multitude of body systems: in facial and skeletal muscles, in the autonomic nervous system (regulating heartbeat, sweating, breathing, digestion), and in the endocrine system (regulating adrenalin, for example). The central idea of the Jamesian theory is that these changes are rich enough to distinguish through the resulting sensations different types of emotions broadly corresponding to distinctions encountered in ordinary language (Deonna and Teroni, 2017). Whether one considers this program of emotion individuation through profiles of bodily sensations promising or not, the fact remains that these sensations do not seem to directly affect the happiness of the subject experiencing them. The challenge posed by the ascetic takes its full significance here: if any sensation can leave us cold—that is, can fail to satisfy the nonalienation constraint—then this is true for the configurations of bodily sensations recruited by James. This means that the Jamesian theory cannot satisfy the intuition that emotions are, as such, hedonic constituents of happiness.⁴

The way in which we have just used the distinction between bodily sensations, on the one hand, and the fact that a subject takes pleasure in them, on the other hand, points in the direction of a conception of the emotions that would make it possible to avoid the difficulties faced by the Jamesian theory. The latter is indeed often criticized for its inability to account for the intentionality of emotions: bodily sensations are at best about bodily changes, whereas fear is about the approaching dog; anger, about a colleague's remark; amusement, about a joke, and so on (of course emotions can sometimes be about sensations, as when we are looking forward to a forthcoming caress). This is why emotions should rather correspond to the idea of "taking (dis)pleasure," which we have just seen at work in the criticism of the Jamesian theory: sensations, as well as reading a novel, listening to an opera, and all other things, should move us positively or negatively in order to contribute to our happiness. According to a popular approach, we should analyze the phenomenon of "taking (dis)pleasure" by means of evaluative representations. To take pleasure in something is to represent it positively; to take displeasure in it is to represent it negatively. In contemporary philosophy of emotions, the most influential theory of this type conceives emotions as perceptual experiences of specific "thick" values.⁵ To be frightened by a dog is to have a perceptual experience of its dangerousness; to get angry at a remark is to have an experience of its offensiveness; to be amused by the joke is to have an experience of its comic character (Tappolet, 2000, 2016; Döring, 2007).

Do we have here the necessary ingredients to support the claim that emotions engage us as such and thus to satisfy the nonalienation constraint? At first sight, yes: speaking of value representation or evaluation seems to implicate the subject in the relevant way. On closer inspection, however, representing the positive or negative value of an object is not yet being committed in any way. Consider the perceptual experiences on the model of which the theory in question understands emotions—for example, the visual experience of a pink poppy. Such an experience may fail to engage the subject in any way; at most it can be said to involve an "impression of truth or presence" of that poppy (Tucker, 2013). The same is true of perceptual experiences regardless of the properties they represent, and this includes the thick values represented by emotions according to the theory at issue. Indeed, it is not clear why the mere fact of representing a remark as offensive would transform the characteristic relation the subject has to his or her environment in perceptual experience into a more substantial kind of engagement. The nonalienation constraint is thus still not met, and hence the intuition that emotions as such contribute to happiness cannot be satisfied by a perceptual theory of the emotions.

The lesson to be drawn is that "taking (dis)pleasure in" cannot be reduced to perceptual experiences of values. In order to find the kind of engagement required, we must try to substitute the psychological mode of "perceiving" or "representing" at the heart of representational theories with another mode that can do justice to the intimate links between emotions and actions, links that are already suggested by the very term "emotion."

A number of theories of the emotions satisfy the nonalienation constraint by insisting on the close links between emotions, motivation, and action.⁶ We shall focus here on the attitudinal theory (Deonna and Teroni, 2012, 2015, 2022), which seems to us to illustrate particularly well the type of engagement suited to meet this constraint. How does this theory understand the emotions? Where representational theories, such as the perceptual theory, appeal to only one psychological mode, that of perceiving or representing, directed towards a content in which the relevant thick values feature, the attitudinal theory redistributes the cards: different types of emotions are so many distinct evaluative modes directed towards contents in which the relevant thick values do not necessarily feature. The perceptual theory asserts that to be angry at a remark is to have the perceptual experience of this remark's offensiveness. The attitudinal theory, on the other hand, asserts that to be angry is to be in a psychological mode specific to anger directed towards that same remark. How to characterize this psychological mode? While various options are open, the version of the attitudinal theory that we prefer conceives of these psychological modes as specific types of engagement with the relevant objects in a way that is directly relevant to our purpose. Specifically, the theory appeals to feelings of action readiness directed toward some aspects of the environment (Deonna and Teroni, 2012, 2015).

The central idea is that we grasp the sense in which emotions are evaluative attitudes when we realize that they are experiences of the body disposed or ready to act in a differentiated way towards a given object or situation. These felt bodily attitudes are typically directed towards objects outside the body of the subject who feels them. Moreover, the idea of attitude must be understood in a broad sense to include tendencies to move away from, to approach, or to go against an object, to focus on it, to submit to it, to see oneself attracted by it, to disengage from it, or even to suspend all interaction with it. Consider anger. In an episode of anger, we feel our bodies prepare for a form of active hostility. This idea is happily applied to other types of emotions. When we are afraid, we feel our bodies mobilize to neutralize something; when we are ashamed, we feel a readiness to move away from the gaze of the person who is causing this emotion; and when we are sad, we feel symptoms of withdrawal from the object whose loss is causing this emotion. Finally, it is important to appreciate that these emotional attitudes harness the subject's attentional resources and direct them to prioritize processing of the situation towards which the emotion is directed (Scarantino, 2014).

Through these examples, which all underline the agential dimension of the feelings characteristic of emotions, we can appreciate how they are evaluative attitudes. Having such an attitude towards a given object is subject to a norm of appropriateness or correctness referring to the relevant thick value (D'Arms and Jacobson, 2000). The attitude in anger—namely, a bodily preparation for hostility—is an attitude that it is appropriate or correct to have toward a given person if and only if he or she has committed an offence. Fear of a dog is another evaluative attitude, which consists in the experience of the body preparing to limit an object's impact (flight, preemptive attack, immobility, etc.), an attitude that it is

correct to adopt if and only if that object is dangerous. And shame is an experience of self-degradation, insofar as we feel our bodies ready to disappear underground or out of sight of those who denigrate us, a correct attitude when we are indeed degraded. These are just a few examples of an approach to emotions that seems attractive to us—its fruitfulness depends of course on the possibility of describing all emotions in a rich and convincing way in terms of felt bodily attitudes.⁷

If we adopt this attitudinal approach to the emotions, the immediate benefit for our purposes is that they possess the right profile to satisfy the nonalienation constraint and, hence, to be hedonic constituents of the happiness of the subject who feels them. Indeed, the types of preparation to act recruited by the attitudinal theory are attitudes that the subject takes towards what is represented. According to this theory, the subject undergoing an emotion does not remain cold in relation to what it represents, be it the ascent of a mountain pass, the experiencing of an opera, the tasting of a chocolate bar, the experiencing of a bodily sensation, and so on. We thus escape the difficulties raised by both the Jamesian and perceptual theories.

Is this really the case? Doesn't the challenge posed by the ascetic resurface in a slightly different way, but with the same unfortunate consequences, for emotions conceived in line with the attitudinal theory? Consider a situation in which you enjoy a chocolate bar, but are, because of your diet, angry with yourself and feel slightly guilty. Or consider a case of morbid curiosity in which you feel a great disgust for an object yet are interested in exploring it further. In such situations, should we not conclude that the emotion of delight or disgust itself does not engage you, your engagement being manifested instead in your anger, guilt, or interest? Doesn't this show that subjects' engagement in their emotions is always optional? No, these verdicts would be too hasty. What these situations suggest is at most that the type of implication of the subject characteristic of the emotions is, under certain conditions, opposed or even cancelled. The cases we have just presented do not call into question the idea that we are engaged in the joy of tasting the chocolate bar or in the disgust that an object arouses in us. Rather, they call into question the kind of engagement on our part that these emotions manifest in complex situations involving other attitudes and, in particular, involving emotions that are in tension with them.

How to deal with these relatively common cases in our investigation of the contribution of emotions to happiness? There is no obvious answer, but here are a few thoughts. When a given situation arouses two contradictory emotions, as in morbid curiosity, then their respective contributions will remain distinct: the contribution of curiosity will be that of a pleasure; that of disgust, a displeasure.⁸ Things get more complicated when the situation features “embedded” emotions, as when joy leads to anger or guilt. Here, two options are available. First, one might go for a “winner takes all” strategy and say that the higher-order emotion fixes the only contribution of the emotional set at issue to happiness. If your joy makes you angry, this is a displeasure that contributes negatively to your

happiness. Alternatively, we can apply the recipe suggested for morbid curiosity and say that, in such a situation, we are engaged in two different ways, each making its own contribution to our happiness: the pleasure of rejoicing contributes positively; the displeasure of anger and guilt, negatively. Whether we choose one or the other of these treatments of complex emotional situations, they do not call into question the attitudinal approach recommended here.

In this section, we have seen how an attitudinal approach to the emotions can meet the nonalienation constraint and, hence, stand by the intuition that emotions are hedonic components of happiness. However, by admitting that emotions are distinct attitudes of preparation for action, have we not called into question the unity characteristic of states of (dis)pleasure? It is to this question that we now turn.

§2 A UNIFIED CONCEPTION OF PLEASURES

In the hedonistic perspective that we have endorsed from the beginning of our discussion, mental states contribute to happiness in that they are (dis)pleasures. However, exploration of the nonalienation constraint has led us to adopt a specific type of attitudinal theory. And it is not easy to see how, on this theory, emotions qualify as pleasures and displeasures. Let's take a step back.

Traditionally, a (dis)pleasure is a hedonic *quale* that is often conceived as a primitive given. Understood in this way, all (dis)pleasures are constituted by *qualia* of this nature—even if these *qualia* can of course be combined with other aspects of the experience of the situation or activity that generates the subject's pleasure. This conception is often criticized for not being able to respect the heterogeneity of (dis)pleasures. Compare the pleasures of savouring chocolate, listening to opera, reading a novel, riding a bike up a mountain, and winning a competition (Clark, 2005). The conclusion that many draw is that such a comparison does not support the idea that these different pleasures are pleasures because they all contain the same hedonic ingredient. In light of this, the strategy that is often seen as the most promising is to admit a great phenomenological diversity in the experiences or, more broadly, situations or activities that give us (dis)pleasure, and to explain what makes them (dis)pleasures by claiming that they are all accompanied by an attitude of taking pleasure or displeasure in the relevant experiences, situations, or activities (Feldman, 2004, but see Bramble, 2013 and Labukt, 2012).

The viability of this strategy must, of course, be measured by what we can say about the nature of these two attitudes. If, for example, we were led to appeal once again to irreducible hedonic *qualia* at the level of these two attitudes, then we would again encounter the problem of heterogeneity. In fact, the most influential advocate of this approach, Fred Feldman, does not say much about the nature of these attitudes, and he even insists that they do not necessarily have any phenomenal properties. The claim that some pleasures are free of any hedonic phenomenology is, it seems to us, an overreaction to the problem of heterogeneity.⁹

An attractive way to navigate these two pitfalls—uniform phenomenology and absence of phenomenology—is to summon the distinction between a determinable and its determinates. A determinable property is a property that an object cannot exemplify as such; an object can exemplify a determinable property only indirectly—that is, by virtue of its exemplification of a determinate property (Johnson, 1921). Colours provide a good illustration of the connection between a determinable property and its determinates. The properties of being coloured and of being red are determinable properties, because an object cannot be coloured or red *tout court*: it is always coloured or red in virtue of the exemplification of a more determinate property—for example, a given shade of vermilion. Having this shade of vermilion is one of many ways of being red or coloured (which is why being vermilion is not the same as being red or coloured), and an object must have a shade in the relevant range to be red or coloured. Moreover, at the level of phenomenology, these links between the determinable “being coloured” and the determinate colours that we perceive allow us to clearly identify the relations between the chromatic properties that we perceive. The perceived colours are unified in that they are determinates of a single determinable, and they are distinct from one another in that they are different determinates. In addition, it should be stressed that the similarities among these various determinates can be arranged in a space structured around three axes that are provided by saturation, tonality, and luminosity.

This distinction between a determinable and its determinates has been applied to (dis)pleasures by Roger Crisp (2006) in order to address the problem of heterogeneity. Regardless of the specific way in which Crisp applies this idea, it is easy to see how it helps avoid the two pitfalls mentioned above: the relation between different (dis)pleasures is not that between different exemplifications of one and the same property, but rather that between different determinates of a given determinable. The way in which this idea should be applied to (dis)pleasures depends of course on their psychological nature. In the foregoing, we have understood emotions as evaluative attitudes that are correct as a function of whether their object exemplifies a given thick value (the threatening, the amusing, etc.). The idea is that these different evaluative attitudes are determinates of two determinables: taking pleasure and taking displeasure. There is not one single attitude of taking pleasure and another single attitude of taking displeasure. Taking pleasure and taking displeasure are *abstractions*. What is psychologically real are the different determinate ways of taking pleasure or displeasure: to have fun, to rejoice, to be proud, to admire, to be amazed, and so on, on the one hand, and to be angry, to be afraid, to be sad, to be ashamed, and so on, on the other hand. The problem of heterogeneity is thus solved, since there are in fact different ways of taking pleasure and displeasure, each with its own phenomenology.

The next question is of course whether these emotional attitudes are not too different at the phenomenological level for it to make sense to put them all under the same umbrella. In the case of colours, this difficulty is avoided by speaking

of (dis)similarities among the perceived colours, (dis)similarities that generate a colour space structured around three axes. What can we say about emotional attitudes if we conceive them as action tendencies? Here, too, different dimensions of (dis)similarity can be highlighted within the relevant experiences. The degree of bodily arousal or relaxation, approach or avoidance, and attentional involvement (local or global, intrinsic or instrumental), and the sense of dominance or submission are all dimensions along which different emotional attitudes may be more or less similar. For example, the characteristic attitudes of fear and anger are similar in that they both involve a high degree of bodily arousal and sustained attention to their object, but they differ in that anger is characterized by a sense of dominating the situation; fear, by a sense of being dominated by it. We can also look at the phenomenology of these attitudes from a more local perspective, asking what kinds of bodily sensations they include (muscle tension, heart rate, localization of sensations in this or that part of the body, etc.) (Deonna and Teroni, 2017, 2020; Numenmaa et al., 2014). Thus, two emotional attitudes can also be compared in light of these more localized sensations.

Furthermore, the different emotional attitudes constitute determinates of the determinables “taking pleasure” and “taking displeasure” in that they constitute different ways of accepting or rejecting a given value. An emotional attitude welcomes a value when it favours, promotes, supports, or protects it. It rejects a value when it neglects it, avoids it, or suppresses it. As far as emotional attitudes can be classified in these ways, they constitute determinates of these two determinables. This leaves open the possibility of admitting, as we shall see in the next section, the existence of borderline cases.

If we subscribe to such an approach to (dis)pleasures in terms of evaluative attitudes, what are its consequences for the question of the respective contributions of different pleasures and displeasures to happiness? One of the promises of traditional hedonism, which we also find in the attitudinal hedonism defended by Feldman, is the possibility of quantifying (dis)pleasures, of comparing them to each other, and of calculating happiness by adding up pleasures and subtracting displeasures. This promise seems to presuppose an approach in terms of hedonic *qualia*, or in terms of two attitudes that can vary in intensity and duration: one of pleasure and another one of displeasure. This approach is much more difficult to defend when the attitudes of taking pleasure and taking displeasure are conceived as determinables that are psychologically realized by determinate attitudes such as enjoying, being amused, being afraid, or being disgusted. How indeed can we go about comparing an attitude of rejoicing at degree n for fifteen seconds with an attitude of feeling pride at degree n^* for eight seconds—for example, in order to determine their respective contributions to happiness? The heterogeneity of the relevant attitudes, as they are portrayed in the approach we have put forward, suggests that such comparisons are at best difficult, if not meaningless. Is this a serious problem? Clearly, the implication of our approach is that it is difficult to quantify how a rich and varied emotional life contributes to happiness. This is a problem if one considers that the quantitative ideal promoted by traditional hedonism is appropriate to the domain. However one wishes

to position oneself in relation to this issue, it seems to us that our approach is exactly in line with the doubtful or even dazed reaction that many of us have to attempts to measure the respective impact of different emotions on happiness.

In light of what we have seen in this section, the attitudinal approach to emotions seems to navigate the two pitfalls of adopting a uniform hedonic phenomenology and of claiming that the nature of (dis)pleasures is not to be located in their phenomenology: emotional attitudes possess a rich phenomenology that allows us to anchor our intuitions concerning both the unity and the diversity of the hedonic domain.

§3 ON ATTITUDINAL HEDONISM

In discussing the contribution of emotions to well-being, we have emphasized the advantages of satisfying constraints of nonalienation and unity, and we have sought to show that an attitudinal approach to emotions can meet these constraints. Yet we may think that this approach encounters difficulties related to the way in which it accounts for the hedonist intuitions from which we have started. Doesn't adopting an attitudinal approach amount to abandoning *de facto* the hedonistic perspective? This difficulty comes in the form of necessity or sufficiency. On the one hand, one can ask whether there is really a *necessary* link between the hedonic components of happiness and tendencies to act. On the other hand, one may ask whether a tendency to act is *sufficient* in itself to make a hedonic contribution to happiness. This last section focuses on these two worries.

Are the hedonic components of happiness *necessarily* action tendencies? Have we not given up on our intuitions about the (un)pleasantness of emotions by conceiving them as psychological states constituted by felt tendencies to act? Two types of mental states—some bodily sensations and moods—call into question the necessity of such a relation, since they intuitively appear to be important hedonic components of happiness. What consequence should we draw from this? Should we abandon the necessity of the relationship and thus a unified approach to the hedonic components of happiness, or should we seek to apply an attitudinal approach to the sensations and moods at issue?

As far as *bodily sensations* are concerned, the worry may seem serious. Indeed, do we really want to say that all bodily sensations are (dis)pleasures because of the presence of an evaluative attitude, rather than because of their exemplification of a hedonic *quale*? The intuition that we should go for the second option is particularly strong with respect to pain sensations. Yet there is good reason to think that the lesson we have drawn at the beginning of our discussion concerning the “pleasures” of the ascetic applies equally to the “displeasures” of pain sensations. In both cases, we must avoid returning to the idea that there is a positive or negative hedonic *quale* shared by all (dis)pleasures. The fact that this is true even for pain sensations can be seen in the phenomenon of pain asymbolia, a dissociative state well documented by clinical studies (see, for example, Grahek, 2007). What is it?

Asymbolic patients can refer to and recognize their pains, but they are not moved by them at all. When an experimenter pierces their skin with a needle, they tend to smile and say something like “It hurts, but I don’t care.” In our view, we should understand pain asymbolia as follows: asymbolic patients refer to distinct bodily sensations that happen not to displease them (Grahek, 2007; Teroni, 2019). Given that asymbolic patients describe their sensations as painful, one might initially think that this conception of their condition is inconsistent with the evidence. When they say that their sensations are painful, doesn’t that mean that they experience the archetype of displeasure? And, since this displeasure is not accompanied by any negative attitude, should we not abandon our attitudinal approach in favour of a more traditional strategy in terms of hedonic *qualia*? No. In fact, pain sensations differ from other sensations in that they represent bodily harm (e.g., Tye, 1995). This opens the possibility that asymbolic patients use the term “painful” to describe bodily sensations that represent damage. The peculiarity of their condition is that their pains (i.e., sensations that represent parts of their body as damaged) are not displeasures. What is disconcerting about asymbolic patients is that their proprioceptive representations of bodily damage (i.e., their pains) always “leave them cold.” That is, their pain sensations are not accompanied by the kind of evaluative attitude of rejection that is familiar to us and that makes them displeasures. The idea is that the representation of bodily harm specific to pain sensations does not explain what makes them displeasures—what explains it is rather the presence of an attitude of rejection toward them. This allows for a unified approach to unpleasant bodily sensations. Many other unpleasant sensations are after all not painful—that is, they do not represent bodily harm (tickling, itching, etc.). What makes all these sensations unpleasant is that they are accompanied by attitudes of rejection. This is typically the case for most of us most of the time, but it is not a necessity. The same strategy applies to “positive” bodily sensations, such as those generated by certain caresses, which are often pleasurable, but may cease to be so.

Let us now consider *moods*. Are moods hedonic components of happiness because they are attitudes? The answer is tricky in that contemporary theories of moods are highly polarized. The most salient dividing line separates theories that make them intentional attitudes whose objects are simply more “global” than those of emotions (Tappolet, 2018) from theories that make them sets of diffuse and enduring bodily sensations (Deonna and Teroni, 2009). Whether one subscribes to one or the other of these theories, we can apply the ideas that we have put forward to moods. If we adopt the first theory, moods, such as anxiety around exam time, should be seen as (dis)pleasures by virtue of the fact that they are rejecting or welcoming attitudes on the model of emotions defended above. “Moody” attitudes will simply be more global than emotional attitudes, and their duration typically longer. If we adopt the second theory, moods—for example, the bodily tension I feel before drinking my first morning coffee—are (dis)pleasures in that they are made up of (dis)pleasurable bodily sensations, which we understand on the model sketched above. In either case, it is not obvious that moods are a serious threat to the approach we advocate.

We now turn to the worry regarding *sufficiency*: is every felt tendency to act really a hedonic component of happiness? After all, some felt tendencies do not seem to contribute to happiness. Consider, for example, the carpenter who is about to hit the head of a nail, or the tennis player who is preparing to receive her opponent's serve. The action tendencies in question are clearly felt, but are not (un)pleasant. This is indeed the case, and the approach that we have advocated above is able to account for what singles out the relevant action tendencies. A felt tendency to act is a (dis)pleasure taken in a given situation *only when it is an evaluative attitude*. What does this mean? An evaluative attitude is an attitude that it is correct to adopt with respect to a situation if and only if it exemplifies a certain value. It must be stressed that these correctness conditions are not imposed from the outside but derive from the very nature of these felt preparations to act: it is because fear is a specific attitude of preparation to act that it is correct if and only if there is a threat.¹⁰ It is only when an attitude has such correctness conditions that it makes sense to speak of an evaluative attitude, an attitude that constitutes a (dis)pleasure and thus contributes hedonically to happiness.

Will critics be convinced by this answer? Perhaps not. How, they will insist, does the fact that emotions are evaluative attitudes give them the status of hedonic components of happiness? This residual worry can target all emotions or certain types of emotions only.

If it targets all emotions, the concern is probably that the phenomenology of welcoming or rejecting a situation in a certain way is not enough to account for the fact that emotions are (dis)pleasures. Does our approach have the resources to alleviate this worry? Yes, once we realize that the concern is motivated by disproportionate attention to (un)pleasurable bodily sensations and to their salience within emotions of a certain intensity. For it is undoubtedly the importance given to unpleasant sensations that feeds the idea that the mere presence of an attitude of welcoming or rejecting is not enough to account for hedonic phenomenology. Of course, it must be admitted that, when they are present, these sensations also contribute to (un)happiness. But, in the light of an attitudinal approach to (dis)pleasures, this admission should not lead us astray: as we have observed, the bodily sensations present during emotions are themselves (dis)pleasurable only because of evaluative attitudes directed towards them. And, if emotions are mostly not directed towards sensations, they also are evaluative attitudes and they are (un)pleasant because of this fact. The concern at issue is therefore misleading, because it does not pay sufficient attention to the fact that sensations are (un)pleasant only when they are accompanied by evaluative attitudes. This being admitted, we can maintain that evaluative attitudes directed towards something other than sensations are (dis)pleasurable for the same reason that (dis)pleasurable sensations are.

If we now seek to apply the worry in a more nuanced way only to certain types of emotions, the diagnosis made by an attitudinal approach will be different and will proceed case by case. Let us consider here anger, one of the borderline cases

between taking pleasure and taking displeasure to which we alluded above. One might think that anger, if it is a paradigmatic example of a felt tendency to act with evaluative correctness conditions, is not clearly a component of (un)happiness. The proponent of an attitudinal approach considers that this idea arises not from the fact that anger is a hedonically “silent” attitude, but rather from the fact that it is hedonically ambivalent. Our intuitions about anger are not clear-cut, for episodes of anger—and especially those that can be given free rein—often combine the displeasure of witnessing or remembering an offence with the anticipated pleasure of revenge. Since these two aspects of anger tip it on both sides of the hedonic scale, its overall hedonic contribution to happiness is unclear. If we can apply such a strategy to all the emotions that constitute borderline cases, one may conclude that this is a significant asset of the approach we advocate.

We have argued in this section that an attitudinal approach has the resources to appease the worries related to the necessity and sufficiency of action tendencies to constitute hedonic components of happiness.

CONCLUSION

Starting from the intuition that emotions contribute to happiness, we have explored how hedonism can accommodate this intuition. Through the nonalienation constraint, we realized that this intuition calls for a conception of (dis)pleasures as states that engage the subject in a way that is irreconcilable with the traditional *quale* approach to (dis)pleasures. The conception of emotions that emerged from this discussion equates them with feelings of action tendencies. In light of the unity constraint, we then argued that, while emotions are so many preparations to act, they nevertheless constitute a unified category of mental phenomena consisting in welcoming or rejecting something. Finally, we have responded to some concerns raised by this approach. On the one hand, we have explained how it can be extended to other affective states. On the other hand, we have argued that the role played by action tendencies in our approach to affectivity does not call into question the idea that emotions contribute to happiness because of their hedonic aspect. This is not to say that hedonism has no further and important challenges to meet, or even that we should favour this approach to well-being. Still, we hope to have contributed to recent attempts to make hedonism more attractive.

NOTES

- ¹ Hedonism as a theory of prudential value is a form of monism. As we just observed, in what follows we do not presuppose the truth of monism, but we will simply seek to articulate the connections between hedonism and theories of emotion. For recent defences of hedonism, see Crisp (2006) and Feldman (2004).
- ² One might think that this observation does not apply to some sensations. We will have the opportunity to see later (§3) that it is necessary, even in the case of pain, to distinguish the sensations from what makes them displeasures. In light of this distinction, we will conclude that pain sensations are no more exceptions to this observation than the pleasurable sensations that we are considering here.
- ³ Note in passing that the nonalienation constraint threatens hedonism as much as it threatens the traditional “objective list” theories of happiness. Indeed, it is nowadays admitted that hedonism is a monistic variant of these theories, a variant according to which only pleasure appears on the list.
- ⁴ We are not denying that some configurations of bodily sensations give pleasure to some of us, nor even that some of these configurations give pleasure (almost) universally. We want to stress only that this is not an aspect of these bodily sensations as such.
- ⁵ This term, which originated in Bernard Williams’s (1985) discussion, draws attention to the idea that we just encountered: emotional evaluations are not “thin,” since being offensive or funny are specific ways of being bad or good.
- ⁶ We are thinking here of the “mixed” theory according to which emotions are combinations of beliefs and desires or wishes (Gordon, 1987; Marks, 1982), as well as the motivational theory developed by Scarantino (2014). For discussion, see De Sousa and Scarantino (2021).
- ⁷ For a detailed exploration of how felt action readiness generates bona fide correctness conditions, see Deonna and Teroni (2021 and, especially, 2022). For criticism, see Ballard (2021).
- ⁸ On ambivalent emotions, see Massin (2018).
- ⁹ See Feldman (2004) and, for criticism of this aspect of his approach, Lin (2020) et Mason (2007).
- ¹⁰ We defend this idea in detail in Deonna and Teroni (2015, 2022). For criticisms, see Ballard (2021) and Dokic and Lemaire (2015).

REFERENCES

- Ballard, Brian Scott, "Content and the Fittingness of Emotion", *Philosophical Quarterly*, vol. 71, no. 4, 2021, p. 845-863.
- Bramble, Ben, "The Distinctive Feeling Theory of Pleasure", *Philosophical Studies*, vol. 162, p. 201-217.
- Clark, Austen, "Painfulness Is Not a Quale", in Murat Aydede (ed.), *Pain: New Essays on Its Nature and the Methodology of Its Study*, Cambridge, MA, MIT Press, 2005, p. 177-197.
- Crisp, Roger, *Reasons and the Good*, New York, Oxford University Press, 2006.
- D'Arms, Justin and Daniel Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions", *Philosophical and Phenomenological Research*, vol. 61, no. 1, 2000, p. 65-90.
- Deonna, Julien and Fabrice Teroni, "Taking Explanations to Heart", *Social Science Information*, vol. 48, no. 3, 2009, p. 359-377.
- , *The Emotions: A Philosophical Introduction*, New York, Routledge, 2012.
- , "What Role for Emotions in Well-Being?", *Philosophical Topics*, vol. 41, no. 1, 2013, p. 123-142.
- , "Emotions as Attitudes", *Dialectica*, vol. 69, no. 3, 2015, p. 293-311.
- , "Getting Bodily Feelings into Emotional Experience in the Right Way", *Emotion Review*, vol. 9, no. 1, 2017, p. 55-63.
- , "Emotional Experience: Affective Consciousness and Its Role in Emotion Theory", in Uriah Kriegel (ed.), *Oxford Handbook of Philosophy of Consciousness*, New York, Oxford University Press, 2020, p. 102-123.
- , "Which Attitudes Fit the Fitting Attitude Analysis of Value?" *Theoria*, vol. 87, no. 5, 2021, p. 1099-1122.
- , "Emotions and their Correctness Conditions: A Defense of Attitudinalism", *Erkenntnis*, 2022. DOI: 10.1007/s10670-022-00522-0
- Dokic, Jérôme and Stéphan Lemaire, "Are Emotions Evaluative Modes?", *Dialectica*, vol. 69, no. 3, 2015, p. 271-292.
- Döring, Sabine A., "Seeing What to Do: Affective Perception and Rational Motivation" *Dialectica*, vol. 61, no. 3, 2007, p. 363-394.
- Feldman, Fred, *Pleasure and the Good Life*, Oxford, Oxford University Press, 2004.
- Fletcher, Guy, "A Fresh Start for the Objective-List Theory of Well-Being", *Utilitas*, vol. 25, no. 2, 2013, p. 206-220.
- Gordon, Robert M., *The Structure of Emotions*, Cambridge, Cambridge University Press, 1987.
- Grahek, Nikola, *Feeling Pain and Being in Pain*, Cambridge, MA, MIT Press, 2007.

Haybron, Daniel M., *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*, New York, Oxford University Press, 2008.

James, William, *The Principles of Psychology*, New York, Dover, 1890.

Johnson, W. E., *Logic (Part 1)*, Cambridge, Cambridge University Press, 1921.

Labukt, Ivar, "Hedonic Tone and the Heterogeneity of Pleasure", *Utilitas*, vol. 24, no. 2, 2012, p. 179-99.

Lin, Eden, "Attitudinal and Phenomenological Theories of Pleasure", *Philosophy and Phenomenological Research*, vol. 100, no. 3, 2020, p. 511-524.

Marks, Joel, "A Theory of Emotion", *Philosophical Studies*, vol. 42, no. 1, 1982, p. 227-242.

Mason, Elinor, "The Nature of Pleasure: A Critique of Feldman" *Utilitas*, vol. 19, no. 3, 2007, p. 379-387.

Massin, Olivier, "Bitter Joys and Sweet Sorrows", in Christine Tappolet, Fabrice Teroni and Anita Konzelmann Ziv (eds.), *Shadows of the Soul. Philosophical Perspectives on Negative Emotions*, New York, Routledge, 2018, p. 50-59.

Nummenmaa, Lauri, Enrico Glerean, Riitta Hari and Jari K. Hietanen, "Bodily Maps of Emotions", *Proceedings of the National Academy of Sciences*, vol. 111, no. 2, 2014, p. 646-651.

Prinz, Jesse J., *Gut Reactions. A Perceptual Theory of Emotion*, New York, Oxford University Press, 2004.

Railton, Peter, "Facts and Values", in Peter Railton, *Facts, Values and Norms: Essays toward a Morality of Consequence*, New York, Cambridge University Press, 2003, p. 43-68.

Scarantino, Andrea and Ronald de Sousa, "Emotion", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2018.

Tappolet, Christine, *Émotions et valeurs*, Paris, Presses Universitaires de France, 2000.

———, *Emotions, Values and Agency*, Oxford, Oxford University Press, 2016.

———, "The Metaphysics of Moods", in Hichem Naar and Fabrice Teroni (eds.), *The Ontology of Emotions*, New York, Cambridge University Press, 2018, p. 169-186.

Tappolet, Christine and Mauro Rossi, "Emotions and Wellbeing" *Topoi*, vol. 34, no. 2, 2015, p. 461-474.

Teroni, Fabrice, "Valence, Bodily (dis-)Pleasure and Emotion", in David Bain, Michael Brady and Jennifer Corns (ed.), *Philosophy of Suffering*, New York, Routledge, 2019, p. 103-122.

Tucker, Chris, "Seemings and Justification: An Introduction", in Chris Tucker (ed.), *Seemings and Justification*, New York, Oxford University Press, 2013, p. 1-29.

Tye, Michael, "A Representational Theory of Pains and Their Phenomenal Character", *Philosophical Perspectives*, vol. 9, 1995, p. 223-239.

PEUT-ON ENCORE ÊTRE HÉDONISTE ÉTHIQUE ?

ANTONIN BROI

MEMBRE ASSOCIÉ, SCIENCES, NORMES, DÉMOCRATIE UMR 8011, SORBONNE UNIVERSITÉ

RÉSUMÉ :

L'hédonisme éthique est la théorie selon laquelle les seules choses qui ont de la valeur finale, respectivement positive et négative, sont le plaisir et le déplaisir. Une grande partie du débat philosophique autour de l'hédonisme éthique s'est concentrée sur la nature de ces états affectifs. Cet article se propose de revisiter cette question en examinant les relations entre l'hédonisme éthique et la philosophie du plaisir et de la douleur. L'hédonisme éthique est traditionnellement associé à une théorie qui conçoit le plaisir comme une qualité phénoménale, mais cette théorie est loin de faire l'unanimité. Nous montrerons que la plupart des théories récentes les plus prometteuses en philosophie du plaisir et de la douleur sont en fait incompatibles avec l'hédonisme éthique, ce qui place l'hédoniste dans une position difficile.

ABSTRACT:

Ethical hedonism is the thesis that the only things that have (respectively positive and negative) final value are pleasure and displeasure. A large part of the philosophical debate around ethical hedonism has focused on the nature of these affective states. This article aims at reconsidering this question by looking at the relationships between ethical hedonism and the philosophy of pleasure and pain. While ethical hedonism has traditionally been associated with felt-affect theories of pleasure, these are far from consensual. A closer look reveals that the most promising theories that have been recently developed in these fields seem to undermine ethical hedonism.

INTRODUCTION

Qui n'apprécie pas le plaisir associé à un bon massage, un repas gastronomique ou la pratique de son passe-temps favori ? L'hédonisme, qui tire parti des lieux communs sur la désirabilité des états plaisants pour ériger le plaisir au rang de valeur suprême, a toutefois fait l'objet de critiques féroces. Celles-ci se focalisent tantôt sur les autres choses qui auraient de la valeur, tantôt sur les cas de plaisir et de déplaisir qui n'auraient pas la valeur postulée par l'hédonisme.

Dans cet article, nous nous pencherons sur l'articulation entre l'hédonisme et les théories du (dé)plaisir. Nous chercherons à montrer que les récents développements en philosophie du plaisir et de la douleur, loin d'être neutres vis-à-vis de l'hédonisme, apportent de nouveaux défis.

Cet article abordera l'hédonisme *éthique*, la thèse selon laquelle la seule chose qui est bonne de manière non instrumentale, c'est-à-dire finale¹, est le plaisir et la seule chose qui est mauvaise de manière finale est le déplaisir (ou la souffrance). Autrement dit, selon l'hédonisme éthique le plaisir et le déplaisir sont les seules choses qui ont une *valeur* finale, respectivement positive et négative. « Non instrumentale » ou « finale » signifient ici qu'il ne s'agit pas de quelque chose qui est bon ou mauvais seulement en raison de la valeur des autres choses qu'il permet d'obtenir. L'hédonisme éthique est une thèse axiologique qui se décline en une variante centrée sur le bien-être, auquel cas la valeur en question est une valeur *pour l'individu* ou *prudentielle* – et il s'agit alors d'une théorie du bien-être – et une variante où la valeur en question est *impersonnelle* ou *simplificiter*. Le propos de cet article s'applique aux deux variantes. L'hédonisme éthique diffère de l'hédonisme *psychologique*, la thèse empirique selon laquelle les désirs finaux des individus concernent seulement le plaisir et le déplaisir. Cette thèse ne nous intéressera pas ici. Pour faciliter la lecture, nous parlerons simplement d'hédonisme pour désigner l'hédonisme éthique.

Le débat philosophique autour de l'hédonisme a souvent gravité autour de la nature du (dé)plaisir (voir par exemple les discussions de Feldman, 2004 ; Crisp, 2006 ; Tännsjö, 2007 ; ou Brax, 2009). Ce n'est guère surprenant puisque l'hédonisme est une thèse portant sur le rôle normatif du (dé)plaisir : une bonne compréhension de ce qui est entendu par (dé)plaisir semble donc nécessaire pour évaluer la plausibilité de l'hédonisme.

Le plaisir et la douleur ont tous deux bénéficié d'une attention philosophique plus soutenue depuis le début des années 2000², mais des perspectives différentes sous-tendent les domaines de recherche associés à ces deux concepts. La philosophie du plaisir reste marquée par sa proximité avec des questions éthiques, notamment autour de la valeur du plaisir et des raisons normatives qu'il donne au sujet. De son côté, la philosophie de la douleur est fermement ancrée dans la philosophie de l'esprit, où les préoccupations de naturalisation du mental priment³. La douleur est alors traitée comme un état mental opposant une certaine résistance au physicalisme, la thèse selon laquelle les états mentaux

sont identiques ou réductibles à des états physiques. Ces discussions se sont récemment attachées à expliquer plus spécifiquement le caractère déplaisant de la douleur, dans le cadre de ce que Bain (2017) appelle l'« *affect debate* ». Ces discussions sont particulièrement intéressantes pour le projet du présent article puisque c'est bien le caractère déplaisant de la douleur qui constitue à proprement parler le déplaisir, c'est-à-dire le pendant négatif du plaisir.

Puisque l'hédonisme est une théorie qui porte à la fois sur le plaisir et le déplaisir, ce sont bien évidemment les théories pouvant expliquer les deux qui nous intéressent ici. Il nous faudra donc parfois réinterpréter des théories portant spécifiquement sur l'un ou sur l'autre pour en proposer une généralisation. Dans la suite de l'article et à des fins de simplification, nous parlerons généralement de plaisir au lieu de (dé)plaisir pour éviter les lourdeurs terminologiques.

Dans un premier temps, nous donnerons un aperçu des relations entre hédonisme et théories du plaisir (section 1). Dans les sections 2 à 4, nous présenterons les grandes familles de théories du plaisir, en montrant que seule la première, l'internalisme de la qualité phénoménale, est compatible avec l'hédonisme. Cela pose un problème grave pour l'hédoniste puisque l'internalisme de la qualité phénoménale est souvent considéré comme moins convaincant que les autres théories plus récentes.

1. L'HÉDONISME ET LA NATURE DU PLAISIR

L'hédonisme ne s'accommode pas de n'importe quelle façon de concevoir le plaisir. Pour commencer, l'hédonisme serait profondément fragilisé s'il s'avérait que le plaisir n'existe pas. Une telle position, que l'on pourrait qualifier d'éliminativiste, pourrait prendre sa source dans un scepticisme généralisé à l'égard de la réalité de la conscience, par exemple⁴. Elle pourrait aussi découler de certains doutes sur la démarcation du plaisir : plutôt qu'une entité bien délimitée, on pourrait considérer que le plaisir consiste en une disjonction d'états hétérogènes désignés à tort par le même concept. Cela aboutirait à nier l'unité du plaisir. Hardcastle (1999) a ainsi défendu un éliminativisme à propos du concept ordinaire de douleur. Selon elle, une analyse conceptuelle révèle qu'il doit dénoter une sensation unique et inanalysable. Pourtant, les recherches scientifiques ont établi que nos expériences douloureuses, loin d'être inanalysables, sont en fait composées de divers processus qui peuvent diverger. Notre concept ordinaire de douleur, à strictement parler, ne dénote donc rien.

En philosophie du plaisir, les positions pluralistes, qui conçoivent le plaisir comme relevant de plusieurs types d'états mentaux disparates (discutées en profondeur par Goldstein, 1985), pourraient également amener à penser que le plaisir, tel qu'on le conçoit généralement, n'existe pas. De telles conceptions ont bien évidemment des conséquences radicales pour l'hédonisme. Tous les hédonistes tiennent pour acquis que le plaisir existe bel et bien. S'il n'existe pas, alors accepter l'hédonisme revient à nier qu'il existe des choses de valeur dans notre monde. Cette thèse aurait des implications similaires à celles de l'antiréalisme à l'égard de la valeur, la position selon laquelle la valeur n'existe pas.

Toutefois, quand bien même on tiendrait le plaisir pour un état mental qui existe et qui est bien délimité, il se pourrait encore que notre théorisation du plaisir fragilise l'hédonisme. En effet, celui-ci attribue au plaisir une valeur finale. La possession de valeur finale par le plaisir implique que les individus ont des raisons (normatives) finales de le rechercher. Or, comme nous le constaterons, les différentes théories du plaisir ne s'accommodent pas également bien de ce rôle normatif, loin de là.

Avant de s'attarder sur les détails de ces théories dans les sections suivantes, commençons par un rapide tour d'horizon.

La philosophie du plaisir, historiquement, est marquée par l'opposition entre internalisme et externalisme (suivant la terminologie de Sumner, 1996, p. 87-91). Les théories internalistes assimilent le plaisir à une propriété subjective ressentie par le sujet, autrement dit une qualité phénoménale. Les différentes théories internalistes caractérisent cette qualité phénoménale de différentes manières. On attribue souvent à Moore (1998 [1903]) une théorie du sentiment distinctif (*distinctive feeling*), selon laquelle le plaisir aurait le même statut qu'une qualité sensorielle comme la couleur rouge. La plupart des auteurs, cependant, insistent plutôt sur le caractère évasif et non localisable du plaisir, qui le différencie clairement de qualités comme le rouge (voir notamment Kagan, 1992 ; Crisp, 2006). Le plaisir, en tant qu'il imprègne notre expérience phénoménale, pourrait plutôt être assimilé à un ton hédonique ou une dimension de notre expérience.

Les théories externalistes, quant à elles, sont attitudinales : elles définissent le plaisir en fonction d'une attitude dirigée vers (une partie de) notre expérience consciente. Ce qui rend une expérience *plaisante* est simplement le fait qu'elle est l'objet de cette attitude précise, et elle n'a rien de plaisant en elle-même indépendamment de cette attitude. Puisque la plupart des auteurs postulent que cette attitude est un désir, les théories attitudinales sont généralement des théories basées sur le désir, c'est-à-dire désidératives (voir notamment Alston, 1967 ; Carson, 2000 ; Heathwood, 2007).

Plus récemment, deux autres familles de théories, développées avant tout en philosophie de la douleur, sont venues enrichir les termes du débat. Alors que les théories précédentes basées sur le désir étaient externalistes, plusieurs auteurs ont proposé des théories désidératives *internalistes*, où le désir en question, plutôt que d'être extérieur à l'expérience plaisante, en fait pleinement partie (Heathwood, 2006 ; Brady, 2018 ; Jacobson, 2018). Dans ce cas, l'expérience plaisante est conçue comme la combinaison d'un désir et d'un autre état qui peut être interprété comme une condition de satisfaction du désir (c'est-à-dire la condition qui réalise le désir). Autrement dit, le plaisir n'est rien de plus qu'un (certain type de) désir satisfait.

D'autre part, les théories dites représentationnalistes ont fait récemment l'objet d'intenses débats en philosophie de la douleur (voir Cutter et Tye, 2011, 2014 ; Bain, 2013, 2017 ; Boswell, 2016 ; Carruthers, 2018). Ces théories conçoivent

le plaisir comme une représentation de valeur, plus précisément comme la représentation d'un état de notre corps ou de notre environnement comme étant bon pour le sujet. De la même manière qu'une perception nous informe sur notre environnement non mental, et peut être véridique ou non selon que son contenu est vrai ou faux, le plaisir nous informe sur la valeur des choses qui nous entourent. La plupart des philosophes s'accordent à dire que le plaisir a un rôle représentationnel. La spécificité des théories représentationnalistes, toutefois, est d'assimiler entièrement la nature du plaisir à ce rôle. Ces théories s'inscrivent généralement dans un programme de recherche plus large, celui du représentationnalisme (fort), visant à rendre compte de tous nos états mentaux, ainsi que de leur caractère phénoménal, en considérant leur contenu représentationnel.

Comment ces théories s'articulent-elles avec l'hédonisme ? Dans les sections qui suivent, nous montrerons que, hormis les théories de la qualité phénoménale, les trois autres familles de théories, chacune à sa manière, mettent en difficulté l'hédonisme.

2. INTERNALISME ET EXTERNALISME À PROPOS DU PLAISIR

Il est généralement reconnu que l'internalisme, dans sa variante standard de la qualité phénoménale, est particulièrement favorable à l'hédonisme, tandis que les théories externalistes attitudinales lui sont défavorables. C'est ce qui explique que le débat entre internalisme et externalisme se soit souvent directement superposé au débat sur le statut normatif du plaisir.

Les théories internalistes de la qualité phénoménale, en effet, postulent que le plaisir n'est rien de plus qu'une certaine qualité phénoménale. Bien qu'il soit en principe possible d'assimiler le plaisir à des qualités phénoménales variées, la plupart des philosophes ont postulé que le plaisir était une qualité phénoménale *sui generis*, c'est-à-dire qui forme sa propre catégorie, séparée des autres types de qualités phénoménales. Considérons par exemple l'expérience plaisante de dégustation d'un plat délicieux. À côté des qualités sensorielles associées au plat (son odeur, son goût, sa texture en bouche), on trouverait la qualité phénoménale caractéristique du plaisir. Il en irait de même pour les autres expériences plaisantes. Se réjouir du succès scolaire de notre enfant, ce serait, d'un côté, penser à son succès scolaire, de l'autre, éprouver cette qualité phénoménale plaisante. Beaucoup de philosophes et de scientifiques ont décrit cette qualité phénoménale comme étant simplement bonne ou comme « *feeling good* » (voir par exemple Katz, 1986 ; Crisp, 2006 ; Frijda, 2009 ; Smuts, 2011). Cela suggère que l'hédoniste peut aisément lui attribuer les propriétés normatives souhaitées en souscrivant à la thèse suivante :

Statut normatif du plaisir₁ (SNP₁) : le plaisir a une valeur finale en vertu de son caractère phénoménal spécifique.

L'idée selon laquelle un caractère phénoménal pourrait donner lieu à une valeur finale est contestable. Il n'en reste pas moins que, si cette idée est acceptée, l'in-

ternalisme de la qualité phénoménale semble fournir une fondation particulièrement solide sur laquelle l'hédonisme peut s'appuyer. Malheureusement, cette conception du plaisir a été rejetée par un grand nombre de philosophes. On peut donc s'interroger sur la compatibilité avec l'hédonisme des autres conceptions qui lui ont été préférées.

Deux arguments en particulier ont incité la plupart des philosophes à s'éloigner de l'internalisme de la qualité phénoménale. Le premier est le problème de l'hétérogénéité des plaisirs. L'argument est déjà présent dans les écrits de Henry Sidgwick. Dans un passage célèbre des *Methods of Ethics*, il soutient que toutes les expériences plaisantes sont désirables :

Devrons-nous alors dire qu'il y a une qualité ressentie mesurable exprimée par le mot « plaisir », qui est indépendante de sa relation à la volition, et strictement indéfinissable dans sa simplicité ? – comme la qualité ressentie exprimée par « sucré » dont nous sommes conscients à des degrés d'intensité variables. Cela semble être la vision de certains auteurs : mais, pour ma part, quand je réfléchis à la notion de plaisir – en utilisant ce terme dans son sens inclusif que j'ai adopté afin d'inclure la gratification intellectuelle et émotionnelle plus raffinée et subtile aussi bien que les plaisirs sensoriels, qu'ils soient plutôt vulgaires ou raffinés – la seule qualité en commun que je peux trouver semble être cette relation au désir et à la volition exprimée par le terme général de « désirable⁵. » (1981 [1907], chapitre II, p. 127⁶)

Il faut noter que l'utilisation du terme « désirable », qui semble encore impliquer la possession par le plaisir d'une certaine propriété intrinsèque – le plaisir serait *digne d'être désiré* –, rend la position de Sidgwick moins limpide que celle de ses héritiers.

En plus du problème de l'hétérogénéité des plaisirs, les théories internalistes de la qualité phénoménale semblent laisser ouverte la possibilité que l'individu ait une attitude négative à l'égard de son plaisir. En effet, si le plaisir ne consiste qu'en une certaine qualité phénoménale, il ne paraît pas en mesure d'imposer un type d'attitudes à son égard : on pourrait donc envisager des individus qui aient une aversion au plaisir, ce qui semble particulièrement contre-intuitif. Autrement dit, si le plaisir n'est rien de plus qu'une qualité phénoménale, à l'instar de la couleur rouge ou du volume sonore, il est difficile d'expliquer que l'on tend à le désirer et à le rechercher. Le succès de cette objection repose crucialement sur l'idée selon laquelle une qualité phénoménale, en tant que qualia, n'a pas de pouvoir causal et par conséquent ne peut pas systématiquement causer certaines attitudes chez le sujet.

Pour expliquer les désirs à l'égard du plaisir et proscrire la possibilité d'un individu qui ne serait pas motivé du tout par ses expériences plaisantes, une option possible est d'intégrer le désir dans la définition même du plaisir. C'est ce que fait précisément l'externalisme attitudinal, qui permet donc de répondre aux

deux principales objections à l'encontre de l'internalisme de la qualité phénoménale. Voici par exemple la théorie proposée par Alston (1967) : « obtenir du plaisir est faire une expérience que l'on préférerait, au moment où on l'a, avoir plutôt que ne pas avoir, sur la base de sa qualité ressentie, indépendamment de toute considération supplémentaire concernant ses conséquences⁷ » (p. 345). À noter que cette définition, comme celles des autres auteurs externalistes, exclut les cas où notre attitude serait purement instrumentale. Heathwood (2007) propose sans doute la variante la plus sophistiquée de cette théorie dans le cas des plaisirs sensoriels : « une sensation S, ayant lieu à l'instant t, est un plaisir sensoriel si et seulement si le sujet de S désire, de manière intrinsèque et *de re*, à l'instant t, que S ait lieu à l'instant t⁸ » (2007, p. 32) Que l'expérience soit désirée de manière *de re* revient à dire que l'objet du désir est exactement la sensation qu'on a au moment t, et non le souvenir de cette sensation, par exemple. Dans le cas précédent du plaisir gustatif, il n'y aurait donc pas de qualité phénoménale plaisante côtoyant les qualités sensorielles associées au plat, mais plutôt un désir dont l'objet serait l'ensemble (ou une partie) de ces qualités sensorielles, en vertu duquel ces dernières seraient plaisantes.

Qu'advient-il alors du statut normatif du plaisir ? Puisque le plaisir n'est plus une qualité phénoménale, SNP₁ doit être abandonné, mais le recours à un désir pour définir le plaisir offre une nouvelle option : c'est le fait que le plaisir est désiré de manière finale qui le rendrait bon de manière finale. L'externalisme attitudinal peut donc aisément s'accompagner de la thèse suivante :

Statut normatif du plaisir₂ (SNP₂) : le plaisir a une valeur finale en vertu du fait qu'il est désiré de manière finale.

La plausibilité de SNP₂ repose sur l'acceptation de la thèse plus générale selon laquelle la valeur finale qu'une chose possède et nos raisons normatives finales à son égard sont déterminées par nos désirs à son égard. On reconnaît ici une thèse subjectiviste d'inspiration humienne à propos de la valeur et des raisons normatives. Les choses se corsent alors pour l'hédoniste : même si ces théories peuvent rendre compte de la valeur finale du plaisir, elles peuvent difficilement rendre compte de la *spécificité normative* du plaisir. En effet, on peut raisonnablement penser que de nombreuses choses en dehors du plaisir sont l'objet de désirs finaux, par exemple la réussite professionnelle, le bien-être de nos proches, la paix dans le monde, la connaissance, la sagesse, etc. Nier ce constat (et maintenir que seul le plaisir fait l'objet d'un désir final) reviendrait à soutenir un hédonisme psychologique plus que douteux. Si l'on accepte la thèse subjectiviste selon laquelle la valeur de quelque chose est déterminée par nos désirs à son égard, on doit alors conclure que de nombreuses choses en dehors du plaisir ont de la valeur finale, ce qui invalide l'hédonisme. Autrement dit, accepter SNP₂ dépouille le plaisir de sa spécificité normative et sape l'hédonisme en profondeur⁹.

On comprend mieux pourquoi le débat entre internalisme et externalisme est d'une importance capitale pour les hédonistes et leurs critiques, au point qu'il

s'entremêle complètement avec la question du statut normatif du plaisir¹⁰. Alors que l'internalisme de la qualité phénoménale semble particulièrement compatible avec l'hédonisme, l'externalisme désidératif tend à nous entraîner vers un subjectivisme qui le contredit.

3. LES THÉORIES INTERNALISTES ATTITUDINALES

L'opposition simple entre l'internalisme de la qualité phénoménale et l'externalisme désidératif a été remise en question par le développement de nouvelles théories du plaisir, qui viennent changer la donne.

Les théories *internalistes* attitudinales, qui placent l'attitude au sein même de l'expérience plaisante, ont ainsi pour but de répondre aux difficultés auxquelles est confronté l'externalisme désidératif. Un des principaux objectifs de ces théories est de pouvoir répondre à l'objection dite d'Euthyphron contre les théories externalistes. Ces dernières soutiennent qu'une expérience est plaisante parce que le sujet a une certaine attitude à son égard. Or il semble plus intuitif de penser que c'est *parce qu'*une expérience est plaisante que le sujet a une certaine attitude à son égard. L'objection d'Euthyphron souligne donc que l'ordre d'explication fournie par les théories externalistes est incorrect.

Une solution séduisante est d'internaliser l'attitude postulée par les théories externalistes. L'idée est alors que le plaisir n'est pas (seulement) l'objet d'une certaine attitude, mais qu'il est constitué par une attitude (en combinaison avec d'autres états mentaux).

C'est une manière de comprendre la théorie du plaisir proposée par Parfit (2011), qui parle d'appréciations hédoniques (*hedonic likings*) pour désigner les attitudes a-rationnelles que l'on a à l'égard de certaines sensations quand on a une expérience plaisante. Or, selon Parfit ce n'est pas la sensation elle-même qui nous donne des raisons normatives d'agir à son égard, mais l'état composite de la sensation et de son appréciation. Le caractère plaisant n'est donc pas une propriété intrinsèque de ces sensations, car il dépend de la présence de l'appréciation. Il peut donc être compris comme survenant sur leur combinaison, ce qui revient à dire que l'appréciation hédonique se situe « à l'intérieur » de l'expérience plaisante. Ainsi, dans le cas de la douleur, « ce qui est mauvais n'est pas notre sensation mais l'état conscient d'avoir une sensation que l'on n'apprécie pas¹¹ » (2011, p. 2)¹². Cette théorie permet de souscrire pleinement à SNP_1 à partir du moment où l'on accepte, comme Parfit, l'idée que l'état composite acquiert un ressenti phénoménologique spécifique en raison de la présence de l'appréciation.

Tournons-nous à présent vers des théories qui ne font pas appel à une appréciation, entendue comme une attitude *sui generis* a-rationnelle, mais plutôt à un véritable désir. Le plaisir peut alors être compris comme un état composite constitué d'un désir et d'un état qui satisfait ce désir. Ce sont donc des théories du plaisir comme *satisfaction d'un désir*. Comme nous le verrons, ces théories

ont des implications normatives intéressantes. Distinguons deux variantes de ces théories, selon qu'elles contraignent le type d'objet du désir en question à des objets *mentaux* ou non.

Une première variante considère que le désir a nécessairement comme objet un état mental. C'est la théorie proposée par Brady (2018), selon laquelle « le caractère douloureux [c.-à-d. déplaisant] ne s'attache pas à la sensation de douleur elle-même, mais est plutôt une propriété de la sensation de douleur plus un désir que la sensation cesse¹³ » (2018). Le déplaisir de la douleur serait donc constitué par une sensation et un désir qu'elle cesse, le désir en question étant donc frustré. On peut aisément étendre cette théorie à l'ensemble du plaisir et du déplaisir.

Une seconde variante admet que le désir peut être tourné vers un objet non mental. Heathwood (2006) défend une théorie de ce type, selon laquelle « S est intrinsèquement content [*pleased*] à t que p ssi S désire intrinsèquement à t que p et S pense à t que p¹⁴ » (2006, p. 557), où le fait d'être « intrinsèquement content » est compris par Heathwood comme un état de plaisir. On retrouve bien un désir envers quelque chose et une condition de satisfaction de ce désir (« S pense à t que p »). Par exemple, l'état de plaisir éprouvé en pensant au succès scolaire de mon enfant s'interprète comme la combinaison de mon désir que mon enfant réussisse à l'école et ma croyance dans son succès scolaire réel. Il faut cependant noter que la condition de satisfaction en question ici diffère de celle en jeu dans la théorie précédente de Brady. Dans cette dernière, l'objet du désir est une sensation et le sujet fait l'expérience de cette sensation au moment où il la désire : le désir est donc *objectivement* satisfait car son objet est effectivement réalisé. En revanche, dans la théorie de Heathwood, la satisfaction en question est subjective car elle consiste en une *croyance* en la réalisation du désir. Un désir peut être objectivement satisfait sans être subjectivement satisfait, et *vice-versa*¹⁵.

En quoi ces théories, dans leurs deux variantes, permettent-elles de répondre à l'objection d'Euthyphron, et donc de justifier notre attitude désidérative et nos raisons normatives finales à l'égard du plaisir ? En concevant le plaisir comme un état de satisfaction d'un désir, elles peuvent avoir recours à l'idée selon laquelle la satisfaction d'un désir est bonne et désirable pour l'individu. Elles peuvent alors souscrire à la thèse suivante :

Statut normatif du plaisir₃ (SNP₃) : le plaisir a une valeur finale parce qu'il est un état de satisfaction d'un désir final.

Arrêtons-nous quelques instants sur cette idée selon laquelle un état de satisfaction d'un désir est bon, car elle est loin d'aller de soi. Cela nous permettra de montrer en quoi SNP₃ diffère de SNP₂.

Supposons que l'on accepte la thèse évoquée dans la section précédente, d'inspiration humienne, selon laquelle on a des raisons normatives de satisfaire les

désirs que l'on a : avoir un désir de X implique que l'on a une raison normative de faire advenir X. C'est la thèse sous-jacente au subjectivisme de SNP₂. Implique-t-elle pour autant que l'on ait une raison normative de faire advenir un désir satisfait, c'est-à-dire la satisfaction d'un désir, la combinaison d'un désir et de son objet ? Pas nécessairement¹⁶. En effet, avoir une raison normative de faire advenir un désir satisfait peut impliquer, selon les circonstances, d'avoir une raison normative de faire advenir *l'objet* d'un désir, si le sujet a déjà ce désir, mais aussi une raison normative de *modifier le désir que le sujet a*, si son objet n'est pas réalisé. Autrement dit, avoir une raison normative de faire advenir un désir satisfait justifie de manipuler non seulement l'objet du désir, mais aussi le désir lui-même. Cela va plus loin que la thèse précédente, car un agent qui se comporterait en accord avec celle-ci pourrait raisonnablement refuser de modifier ses propres désirs. Après tout, c'est *l'objet* de son désir auquel il attribue de la valeur, car le désir ne fournit pas de raison normative envers lui-même ; le désir n'a donc pas de valeur (à moins bien sûr que le sujet ait un autre désir qui aurait comme objet son désir initial).

La thèse selon laquelle on a une raison normative de faire advenir un désir satisfait est donc une extension non triviale de la thèse humienne. Des agents qui ne se comporteraient pas en conformité avec cette extension ne feraient pas preuve d'une irrationalité flagrante. Jacobson la soutient explicitement dans sa défense de sa théorie. Selon elle, « quand on a un désir que C n'ait pas lieu, on n'a pas seulement une raison (de premier ordre) d'éliminer C, mais aussi une raison (de second ordre) d'éliminer l'état complexe qui inclut à la fois C et le fait d'avoir ce désir¹⁷ » (p. 19). Supposons que cette extension de la thèse humienne soit vraie.

Pour juger du succès des théories de la satisfaction de désirs et de leur compatibilité avec l'hédonisme, il est nécessaire de revenir sur la distinction importante entre satisfaction objective et satisfaction subjective de désirs. L'extension de la thèse humienne permet de conclure à la valeur finale de la satisfaction *objective* de nos désirs. Les théories, comme celle de Brady, qui assimilent le plaisir à la satisfaction objective de nos désirs envers d'autres états mentaux occurrents semblent donc à même de défendre avec succès la valeur finale du plaisir. Cependant, il va de soi que l'individu peut aussi avoir, à côté de ses désirs finaux envers d'autres états mentaux, des désirs finaux envers des objets non mentaux, tels que le bien-être de ses proches ou la réussite professionnelle. Or, la satisfaction objective de ces désirs-là ne constituerait pas une expérience plaisante. La théorie de Brady implique donc que d'autres choses que le plaisir ont de la valeur finale. Elle contredit donc l'hédonisme¹⁸.

Qu'en est-il des théories reposant sur la satisfaction *subjective* de nos désirs, l'état composite formé d'un désir et d'une croyance en sa réalisation ? L'extension de la thèse humienne n'implique en rien que cet état composite ait une quelconque valeur finale. En effet, pour qu'il ait une valeur finale, il faudrait que le sujet ait, en plus des raisons normatives finales de modifier son désir, des raisons normative finales de modifier directement sa croyance en la réalisation du désir.

Or ce n'est pas le cas : le sujet a des raisons normatives finales de faire advenir la réalisation objective de son désir, mais il n'a pas de raisons normatives de modifier directement sa croyance en la réalisation du désir. Contrairement à ce que soutient Jacobson, on peut donc avoir des doutes quant à la capacité des théories de la satisfaction subjective des désirs à conclure que le plaisir est bon de manière finale et à défendre avec succès SNP_3 sur la base de l'extension de la thèse humienne. Quoi qu'il en soit, quand bien même ces théories parviendraient à justifier SNP_3 , la valeur finale des désirs *subjectivement* satisfaits dériverait encore de la valeur finale des désirs *objectivement* satisfaits, ces derniers étant distincts du plaisir. L'hédonisme resterait donc incompatible avec de telles théories.

En résumé, il apparaît bien que les théories du plaisir comme satisfaction d'un désir amènent soit à douter de la valeur finale du plaisir, soit à l'accepter mais en acceptant que d'autres choses que le plaisir aient une telle valeur.

4. LE PLAISIR COMME REPRÉSENTATION DE VALEUR

Peut-être trouvera-t-on une fondation plus solide pour l'hédonisme dans les théories représentationnalistes, ou évaluativistes, qui offrent une solution de remplacement séduisante aux théories précédentes. Le plaisir serait alors un état mental qui représente quelque chose comme bon pour le sujet, autrement dit, une représentation de valeur.

Ces théories ont été surtout développées en philosophie de la douleur, où elles reprennent le flambeau des théories perceptualistes comme celles d'Armstrong (1962) et de Pitcher (1970), qui cherchaient à assimiler la douleur à une perception de l'état d'une partie de notre corps (c.-à-d. un certain type de représentation)¹⁹. Cette ambition s'arrêtait cependant à l'aspect déplaisant de la douleur, qui n'était pas assimilé à une représentation. Armstrong (1962) rendait ainsi compte de l'aspect déplaisant en évoquant un désir, ainsi que le font les théories externalistes désidératives de la section 2.

Les théories représentationnalistes développées plus récemment étendent leurs ambitions à l'explication du caractère déplaisant de la douleur. Selon ces théories, une expérience douloureuse est déplaisante en vertu du fait de représenter un dommage corporel *comme mauvais pour le sujet*, comme ayant une valeur négative. Par exemple, selon la théorie évaluativiste de la douleur proposée par Bain (2013, 2017), le caractère phénoménal hédonique peut être assimilé à la propriété d'être bon ou mauvais : « [l]e fait pour un sujet d'être dans un état de douleur déplaisante consiste en (i) le fait d'éprouver une expérience (la douleur) qui représente un désordre d'un certain type, et (ii) la représentation, par la même expérience, du désordre comme mauvais pour lui dans un sens corporel²⁰ » (2013, p. S82). En élargissant cette théorie au pendant positif du déplaisir, on obtiendrait, par exemple, qu'une expérience gustative plaisante attribue la propriété d'être bon à l'ingestion d'un aliment²¹. L'état de plaisir éprouvé en pensant au succès scolaire de mon enfant, quant à lui, pourrait correspondre à

une représentation mentale du succès scolaire de mon enfant comme étant bon. Si l'idée selon laquelle le plaisir a pour fonction de représenter l'état de notre corps ou de notre environnement et d'informer le sujet est globalement consensuelle, il est important de noter que le représentationnalisme, tel que l'ont développé les auteurs précédents, va plus loin – on parle souvent de représentationnalisme fort (*strong representationalism*) pour désigner ces théories. Elles se placent dans la lignée d'un programme de recherche à l'ambition naturaliste qui souhaite réduire le caractère phénoménal de toutes nos expériences à leur contenu représentationnel, en d'autres termes à ce qu'elles représentent. Ce contenu représentationnel, à son tour, serait moins mystérieux et plus aisément réductible à des propriétés compatibles avec le physicalisme, la thèse selon laquelle tous les états mentaux sont identiques ou réductibles à des états physiques. Dans ce cadre, le représentationnalisme fort doit adhérer à la thèse de la transparence de l'expérience²² : en introspectant ses expériences, le sujet n'est censé trouver rien d'autre que ce que son expérience représente, et les propriétés introspectées devraient être exclusivement expérimentées comme des propriétés du monde non mental, jamais comme des propriétés de son expérience. Autrement dit, les expériences sont transparentes : ce sont des fenêtres ouvertes sur le monde non mental. Dans le cas du plaisir, cela signifie que l'introspection du caractère plaisant de notre expérience devrait toujours aboutir à appréhender le caractère bon de quelque chose de non mental, comme le succès scolaire de mon enfant dans l'exemple précédent.

Ces théories peuvent facilement être mobilisées pour soutenir que le plaisir fournit au sujet des raisons normatives d'agir. Si le sujet reçoit l'information selon laquelle quelque chose est bon, il est compréhensible qu'il acquière des raisons normatives de le promouvoir ou de le préserver, et si le sujet reçoit l'information selon laquelle quelque chose est mauvais, des raisons normatives de l'arrêter et de l'éviter²³. Toutefois, si tant est que ces théories permettent bien au plaisir de fournir des raisons normatives finales aux individus, il s'agit évidemment de raisons visant ce qui est représenté comme bon, c'est-à-dire l'objet non mental de notre représentation. Dans le cas d'une douleur déplaisante, qui représente le dommage corporel comme mauvais pour nous, nous acquerions des raisons normatives finales de nous débarrasser du dommage corporel. Par exemple, une douleur au pied à la suite d'une entorse semble nous donner des raisons finales de ne pas nous appuyer sur ce pied, d'en prendre soin, etc. Expliquer avec succès l'existence de ces raisons, qui semblent faire partie intégrante de notre compréhension ordinaire de la douleur, est assurément un atout pour le représentationnalisme, puisque les théories de la qualité phénoménale, par exemple, ne sont pas capables de les expliquer.

Malheureusement, ces raisons ne sont pas les raisons normatives qui intéressent l'hédoniste, car celui-ci postule l'existence de raisons normatives finales *dirigées vers l'expérience plaisante elle-même*. À première vue, les théories représentationnalistes laissent peu de place pour de telles raisons (voir Aydede, 2019). En effet, leur existence contreviendrait au principe de transparence de l'expérience présenté précédemment. Si nous avions effectivement des raisons normatives

finales d'agir à l'égard de notre expérience, cela impliquerait, semblerait-il, que l'on puisse appréhender notre expérience directement, sans passer par ce qu'elle représente.

Cette question n'intéresse pas seulement les partisans et les détracteurs de l'hédonisme. L'existence de raisons normatives finales à l'égard de notre expérience plaisante, au moins dans certaines circonstances, semble évidente pour la plupart des philosophes comme pour le sens commun. Une théorie satisfaisante du plaisir se doit donc d'en rendre compte. En philosophie de la douleur, l'objection du messenger (*messenger objection*) à l'encontre de l'évaluativisme s'attaque précisément à cette conclusion contre-intuitive (voir par exemple Jacobson, 2013). Les théories représentationnalistes ne parviennent pas à expliquer pourquoi quelqu'un pourrait vouloir prendre des médicaments antidouleur. Ces médicaments, en effet, agissent sur l'expérience de douleur et non sur l'état corporel endommagé que cette expérience représenterait. Pourquoi les individus souhaiteraient-ils s'attaquer à cet état affectif, qui, selon la théorie représentationnaliste, ne fait rien de plus que d'informer le sujet, de lui transmettre un message ?

Certains auteurs justifient la prise d'antidouleurs en termes de raisons instrumentales. Cutter et Tye (2011), par exemple, argumentent que l'expérience de douleur déplaisante, indépendamment de la réalité de la douleur représentée, a des conséquences négatives pour le sujet, comme le fait d'être une distraction, d'être dérangement : la prise d'antidouleurs servirait à éviter ces « effets secondaires » de la douleur déplaisante. Cutter et Tye admettent donc que le sujet sait que le dommage corporel que l'expérience déplaisante représente ne disparaîtra pas en faisant disparaître l'expérience. Cette hypothèse est contestée par Simon (2019), dans sa défense de la position qu'il appelle l'évaluativisme naïf faible. Selon cette position, quand les individus s'attaquent à leur état mental en prenant des antidouleurs, ils cherchent bien à faire disparaître le dommage corporel que l'état mental représente. Cela revient à postuler une certaine naïveté de la part des individus, semblable à celle d'un enfant qui fermerait les yeux pour faire disparaître la chose menaçante devant lui.

Quoi qu'il en soit, si l'on accepte ces théories instrumentalistes des raisons normatives d'agir à l'égard du plaisir, il faut se résoudre à admettre la thèse suivante sur son statut normatif, qui est fatale pour l'hédonisme :

Statut normatif du plaisir₄ (SNP₄) : le plaisir n'a pas de valeur finale.

D'autres auteurs justifient la prise d'antidouleurs par des raisons normatives finales, ce qui permet d'écarter la menace que fait peser SNP₄ sur l'hédonisme. Cutter et Tye, dans un autre article (2014), invoquent comme leurs prédécesseurs perceptualistes un désir final dont l'objet serait l'expérience plaisante, que les êtres humains acquerraient normalement au cours de leur développement psychologique. L'existence de ce désir ne relèverait pas d'une nécessité métaphysique, mais simplement d'une régularité attribuable au fonctionnement typique des êtres humains. Si c'est ce désir qui est à l'origine du statut norma-

tif du plaisir, il faudrait accepter SNP_2 , avec les obstacles pour l'hédonisme que cela génère (voir section 2).

Une autre réponse, que Bain (2017) juge particulièrement prometteuse, est de faire appel au caractère *perceptuel* des représentations de valeur en question. Il est généralement reconnu qu'il y a différentes manières de représenter quelque chose, c'est-à-dire différents modes de représentation, par exemple croire, percevoir, espérer, etc. Ces différences sont souvent mobilisées pour expliquer la phénoménologie (ou l'absence de phénoménologie) associée à ces différentes attitudes : une perception ayant le même contenu qu'une croyance s'en distinguerait quand même par son mode de présentation. On pourrait alors avancer que le caractère mauvais du déplaisir serait dû à son mode de présentation spécifiquement perceptuel. Cela permettrait d'expliquer pourquoi le plaisir semble avoir une valeur finale négative alors qu'une croyance ayant le même contenu ne semble pas en avoir. Bain soutient donc la thèse suivante sur le statut normatif du plaisir :

Statut normatif du plaisir₅ (SNP_5) : le plaisir a une valeur finale parce qu'il représente de manière perceptuelle quelque chose comme ayant de la valeur.

L'argumentaire de Bain est contestable (voir Jacobson, 2018). Toutefois, même s'il parvenait à justifier de manière crédible la valeur finale du plaisir, il n'en reste pas moins que celle-ci serait censée dériver de la valeur attribuée à un objet non mental et représentée par le plaisir. À moins de supposer que la représentation de cette valeur soit systématiquement erronée, cela implique qu'il y a d'autres choses qui aient de la valeur finale. Ainsi, si l'on accepte SNP_5 , l'hédonisme devrait de nouveau être rejeté.

Enfin, une dernière réponse est celle de Boswell (2016), qui postule l'existence, dans les états plaisants ordinaires, de deux représentations de valeur distinctes. La première est celle que postulent toutes les théories représentationnalistes et qui représente l'état corporel ou l'état de notre environnement comme bon ou mauvais. La seconde représentation que Boswell met en relief est relative à ce qui est souvent appelé l'« affect secondaire » dans la littérature scientifique. Dans le cas de la douleur, il est généralement reconnu que l'anxiété et l'inquiétude provoquées par l'expérience douloureuse sont sources d'un déplaisir qui vient s'ajouter au déplaisir directement associé à l'expérience douloureuse. Selon Boswell, ce déplaisir secondaire (*secondary unpleasantness*) serait attaché intentionnellement (*intentionally attached*) à la douleur. Le sujet aurait des raisons d'agir à l'égard de son expérience douloureuse en vertu de cette deuxième représentation. En effet, selon Boswell, « [q]uand une expérience émotionnelle déplaisante est attachée intentionnellement à une expérience introspectable e ayant le contenu p, alors (i) quand e n'est *pas introspectée*, l'émotion représente l'occurrence de p comme mauvaise. Et (ii) quand e est *introspectée*, l'émotion représente l'occurrence de e comme mauvaise²⁴ ». La deuxième représentation, durant les phases d'introspection de notre expérience douloureuse, représenterait donc l'expérience douloureuse elle-même comme étant mauvaise. Boswell ne

s'attarde pas sur la justification de ce principe, qui est pourtant loin d'aller de soi puisqu'il semble à première vue contrevenir au principe de transparence que les représentationnalistes sont engagés à respecter. Supposons cependant, aux fins du présent argument, que le principe de Boswell soit valable. Dans ce cas, on aboutirait à la thèse suivante sur le statut normatif du plaisir :

Statut normatif du plaisir₆ (SNP₆) : le plaisir a une valeur finale parce qu'il est représenté par l'individu comme bon.

SNP₆, comme SNP₂, fait reposer la valeur finale du plaisir sur le fait qu'il serait l'objet d'une certaine attitude du sujet – une représentation de valeur pour SNP₆, un désir pour SNP₂ –, ce qui requiert de nouveau de souscrire à un subjectivisme à l'égard de la valeur. Puisque les individus ont des représentations de valeur ayant des objets divers et souvent non mentaux, une telle position débouche également sur la négation de la spécificité normative du plaisir. Son statut est le même que celui de n'importe quel objet représenté comme bon par le sujet.

L'adoption d'une théorie représentationnaliste semble donc être incompatible avec l'hédonisme quelle que soit la manière dont on rend compte de la valeur du plaisir.

CONCLUSION

En somme, les théories développées en philosophie du plaisir et de la douleur donnent du fil à retordre aux partisans de l'hédonisme. Le constat est variable. Pour certaines de ces théories, comme les théories représentationnalistes ou la théorie de la satisfaction subjective des désirs, il semble difficile de rendre compte de la valeur finale du plaisir. Elles sont alors poussées à souscrire à SNP₄, la thèse qui nie la valeur finale du plaisir. Les autres théories peuvent adhérer à SNP₂, SNP₃, SNP₅ ou SNP₆ pour accommoder la valeur finale du plaisir, mais ce faisant elles nient la spécificité normative du plaisir, qui est pourtant requise par l'hédonisme.

On peut tirer de ce constat un argument général à l'encontre de l'hédonisme : il est raisonnable d'exiger de toute théorie éthique qui évoque certaines entités qu'elle soit compatible avec nos meilleures théories de ces entités ; l'hédonisme fait jouer un rôle central au plaisir ; or l'hédonisme n'est pas compatible avec nos meilleures théories du plaisir ; donc l'hédonisme doit être rejeté.

L'hédonisme serait-il donc définitivement mis en échec par les derniers développements en philosophie du plaisir et de la douleur ? Pas nécessairement. On pourrait en fait tout autant renverser l'argument pour s'en servir contre les théories du plaisir incompatibles avec l'hédonisme : l'incompatibilité entre l'hédonisme et certaines théories du plaisir implique soit que l'hédonisme est faux, soit que ces théories du plaisir sont fausses ; or l'hédonisme est vrai ; donc ces théories sont fausses. Cet argument ne convaincra bien sûr que les personnes déjà fermement convaincues par l'hédonisme. Cependant, un argument simi-

laire pourrait être construit en s'appuyant non pas sur l'hédonisme, mais sur la thèse plus faible selon laquelle le plaisir a une valeur finale *sui generis*, c'est-à-dire qui lui est propre. Cette thèse étant incompatible avec les théories récentes du plaisir, l'accepter revient à rejeter ces dernières. La thèse de la valeur finale *sui generis* du plaisir exprime l'idée d'une spécificité normative du plaisir, mais elle est compatible avec l'existence d'autres entités ayant de la valeur finale. On peut donc adhérer à cette thèse tout en rejetant l'hédonisme si l'on pense que l'hédonisme ne pèche pas par son appréhension du plaisir – ce serait bien une entité possédant une valeur finale d'un type spécifique – mais que l'hédonisme va trop loin en excluant toute autre entité du champ de l'axiologie, c'est-à-dire en refusant que d'autres entités (par exemple la connaissance ou la beauté) puissent avoir d'autres types de valeur finale.

Il n'est pas question ici de se prononcer pour une interprétation particulière plutôt qu'une autre. Notre but, plus modestement, était de mettre au jour la manière dont les théories récentes du plaisir viennent contredire l'hédonisme éthique. La prise en compte de ce constat nous paraît être de nature à enrichir non seulement le débat éthique autour de l'hédonisme, mais aussi les discussions en philosophie du plaisir et de la douleur.

REMERCIEMENTS :

Je tiens à remercier les évaluatrices et évaluateurs anonymes de l'article ainsi que les participants à l'atelier « Well-being and Affective States » organisé à l'Université Clermont Auvergne en novembre 2019 pour leurs commentaires très utiles.

NOTES

- ¹ L'hédonisme éthique est souvent formulé en parlant de valeur *intrinsèque* du (dé)plaisir. Malheureusement, il existe différentes manières de comprendre ce que « intrinsèque » signifie (voir Feldman, 1998, pour un passage en revue des différentes interprétations), et l'équivalence valeur intrinsèque = valeur finale n'est qu'une interprétation parmi d'autres.
- ² Pour une vue d'ensemble de la philosophie récente de la douleur, voir Corns (2018) ; pour une vue d'ensemble de la philosophie du plaisir, voir Katz (2016) et, en français, Broi (2019).
- ³ Comment expliquer cette différence de perspective entre la philosophie du plaisir et la philosophie de la douleur ? D'abord, l'hédonisme étant souvent présenté comme une philosophie du plaisir, c'est naturellement sur le plaisir que les éthiciens ont concentré leurs efforts, en présupposant parfois que tous les résultats s'étendaient aisément au pendant négatif du plaisir. Par ailleurs, on pourrait souligner le fait que la douleur est un phénomène souvent plus sensoriel, plus facilement localisable dans le corps, et globalement moins évasif et plus saillant que le plaisir. Elle est de ce fait devenue un cas d'étude classique pour les philosophes s'intéressant aux états mentaux en général, d'où son statut en philosophie de l'esprit.
- ⁴ Dennett (1978, 1999[1986]), par exemple, a défendu une position proche de l'éliminativisme à l'égard de la douleur. Plus récemment, Kammerer (2019) a mis en évidence les conséquences de telles positions sur la valeur du (dé)plaisir.
- ⁵ « Shall we then say that there is a measurable quality of feeling expressed by the word “pleasure,” which is independent of its relation to volition, and strictly undefinable from its simplicity?—like the quality of feeling expressed by “sweet,” of which also we are conscious in varying degrees of intensity. This seems to be the view of some writers: but, for my own part, when I reflect on the notion of pleasure,—using the term in the comprehensive sense which I have adopted, to include the most refined and subtle intellectual and emotional gratifications, no less than the coarser and more definite sensual enjoyments,—the only common quality that I can find in the feelings so designated seems to be that relation to desire and volition expressed by the general term “desirable” »
- ⁶ Toutes les traductions sont les nôtres.
- ⁷ « To get pleasure is to have an experience which, as of the moment, one would rather have than not have, on the basis of its felt quality, apart from any further considerations regarding consequences. »
- ⁸ « a sensation S, occurring at time t, is a sensory pleasure at t iff the subject of S desires, intrinsically and de re, at t, of S, that it be occurring at t »
- ⁹ On retrouve explicitement un tel argument chez Kagan (1992) dans sa discussion de la nature du bien-être. Après avoir présenté la théorie hédoniste, il prend appui sur l'objection de l'hétérogénéité des plaisirs pour adopter une théorie externaliste désidérative selon laquelle « a mental state is pleasant if it is desired » (p. 170). Mais l'hédoniste, avance-t-il, se trouve ici sur une pente glissante. Puisque « [o]ne mental state is more valuable than another — makes a greater contribution to our welfare — by virtue of the fact that we prefer it », on se rend compte que « it is really preference that is doing the work » (p. 171), ce qui nous entraîne irrésistiblement vers la théorie du bien-être subjectiviste selon laquelle ce sont les désirs de l'individu qui déterminent ce qui a de la valeur pour lui. Rachels (2000) évoque le même mouvement argumentatif dans sa discussion de la nature du caractère déplaisant. S'il s'avère

que « unpleasant experiences are non-instrumentally bad due to the fact that one is moved to end them or dislikes them », alors il est probable que « [d]esires confer disvalue or value in other instances » (p. 188).

¹⁰ C'est ce qui explique par exemple qu'Erwin Goldstein (1980), un fervent défenseur de la thèse de la valeur intrinsèque du plaisir, soit amené à la défendre en l'opposant directement à l'externalisme – une thèse sur la nature du plaisir – qu'il décrit comme soutenant que « [t]he attitudes towards the experience enter into the definition of 'pleasure' » (p. 349).

¹¹ « what is bad is not our sensation but our conscious state of having a sensation that we dislike »

¹² Kahane (2009) adhère à une théorie similaire de la douleur, qu'il appelle *Experiential Dislike Theory*. Il parle de souffrance (*suffering*) pour désigner l'état composite de la sensation douloureuse et de la désapprobation du sujet et affirme que « the state that is intrinsically bad is not that of having a sensation of pain, but that of suffering – of having this sensation and disliking it » (p. 333).

¹³ « painfulness does not attach to the pain sensation itself, but instead is a property of the pain sensation plus a desire that this sensation cease »

¹⁴ « S is intrinsically pleased at t that p iff S intrinsically desires at t that p and S believes at t that p »

¹⁵ Jacobson présente une théorie similaire appliquée à la douleur, où le désir postulé a pour objet un état corporel. Selon elle, « painfulness [i.e. unpleasantness] is constituted by a (subjectively) frustrated (first-order) conative (desire-like) attitude that is directed towards the bodily condition the obtaining of which the pain represents » (2018, p. 1). Autrement dit, « pain is constituted by the mental fact of having a frustrated desire » (2018, p. 18), et la frustration en question est subjective.

¹⁶ Voir Bain (2017, p. 7-8) pour une discussion explicite de cette distinction. Ce que nous appelons ici l'extension de la thèse humienne correspond à son *wide-scope Humeanism*.

¹⁷ « when one has a desire that C does not obtain, one not only has a (first-order) reason to eliminate C, but also a (second-order) reason to eliminate the complex condition that includes both C and one's having this desire »

¹⁸ Considérons brièvement une option de remplacement pour maintenir que la satisfaction objective de désirs finaux envers des objets mentaux a une valeur finale que n'a pas la satisfaction objective de désirs finaux envers des objets non mentaux. On pourrait renier SNP₃ et adopter à la place SNP_{3 BIS} : le plaisir a une valeur finale parce qu'il est un état de satisfaction d'un désir final dont le sujet fait l'expérience directe. L'idée serait qu'il y a une relation épistémique très étroite entre le sujet et ses états mentaux conscients – peut-être une relation d'acointance – qui permet au sujet d'avoir accès à leur essence. Le sujet peut ainsi avoir accès à la valeur finale d'un état de plaisir parce que ce dernier est purement mental, ce qui se manifeste par le fait que le plaisir « feels good ». En contraste, le sujet manque cet accès direct à l'état de satisfaction d'un désir dont l'objet est non mental. Il ne vit pas cet état positif « de l'intérieur », pour ainsi dire. Mais cette réponse, quand bien même elle expliquerait de manière convaincante la phénoménologie du plaisir, échoue à démontrer que la satisfaction objective de désirs finaux envers des objets non mentaux n'a pas de valeur finale. En effet, la spécificité à laquelle elle fait appel pour rendre compte de la valeur finale du plaisir – le fait qu'on a directement conscience de l'essence du plaisir dont on fait l'expérience – est une spécificité portant sur notre accès épistémique à sa valeur finale. Ainsi, on ne peut pas en conclure que la satisfaction objective de désirs finaux envers des objets non mentaux n'a pas valeur finale, mais seulement que l'on a un accès moins immédiat, plus indirect, à cette valeur finale.

¹⁹ Voir Aydede (2005) pour une introduction approfondie aux théories perceptualistes.

²⁰ « [a] subject's being in unpleasant pain consists in his (i) undergoing an experience (the pain) that represents a disturbance of a certain sort, and (ii) that same experience additionally representing the disturbance as bad for him in the bodily sense »

²¹ Une autre théorie représentationnaliste est celle récemment défendue par Carruthers à propos de la valence (qu'il assimile au caractère (dé)plaisant de notre expérience), selon laquelle la valence est une « nonconceptual representation of badness or goodness » (2018, p. 6).

- ²² Voir Nida-Rümelin (2007) pour une discussion critique des différentes variantes de la thèse de la transparence. La présentation de la thèse de la transparence qui suit est inspirée de celle d'Aydede (2019).
- ²³ Accepter cela revient à rejeter la théorie humienne des raisons qui veut que les raisons soient fondées (au moins partiellement) sur des désirs, car l'expérience plaisante, selon les théories représentationnalistes, n'est pas un état désidératif et ne comporte pas un état désidératif. Elle est un état similaire à une croyance. On le voit, les différentes théories du plaisir engagent des conceptions radicalement divergentes de la source des raisons normatives et de la motivation morale.
- ²⁴ « When an unpleasant emotional experience is intentionally attached to an introspectable experience *e* with content *p*, then (i) when *e* is unintrospected, the emotion represents *p*'s obtaining as bad. And (ii) when *e* is introspected, the emotion represents *e*'s obtaining as bad. »

BIBLIOGRAPHIE

Alston, William P., « Pleasure », in Paul Edwards (dir.), *The Encyclopedia of Philosophy*, New York, Macmillan, 1967.

Armstrong, D. M., *Bodily Sensations*, Studies in Philosophical Psychology, London, Routledge & Kegan Paul, 1962.

Aydede, Murat, « Introduction: A Critical and Quasi-Historical Essay on Theories of Pain », in Murat Aydede (dir.), *Pain: New Essays on Its Nature and the Methodology of Its Study*, Cambridge, Mass., MIT Press, 2005, p. 1-58.

———, « Is the Experience of Pain Transparent? Introspecting Phenomenal Qualities », *Synthese*, vol. 196, no. 2, 2019, p. 677-708.

Bain, David, « What Makes Pains Unpleasant? », *Philosophical Studies*, vol. 166, no. 1, 2013, p. 69-89.

———, « Why Take Painkillers? », *Noûs*, vol. 53, no. 2, 2017, p. 462-490.

Boswell, Paul, « Making Sense of Unpleasantness: Evaluationism and Shooting the Messenger », *Philosophical Studies*, vol. 173, no. 11, 2016, p. 2969-2992.

Brady, Michael S., « Painfulness, Desire, and the Euthyphro Dilemma », *American Philosophical Quarterly*, vol. 55, no. 3, 2018, p. 239-250.

Brax, David, *Hedonism as the Explanation of Value*, Dissertation, 2009.

Broi, Antonin, « Plaisir (Version Académique) », *L'Encyclopédie Philosophique*, 2020.

Carruthers, Peter, « Valence and Value », *Philosophy and Phenomenological Research*, vol. 97, no. 3, 2018, p. 658-680.

Carson, Thomas L., *Value and the Good Life*, Notre Dame, University of Notre Dame Press, 2000.

Corns, Jennifer, « Recent Work on Pain », *Analysis*, vol. 78, no. 4, 2018, p. 737-753.

Crisp, Roger, « Hedonism Reconsidered », *Philosophy and Phenomenological Research*, vol. 73, no. 3, 2006, p. 619-645.

Cutter, Brian, and Michael Tye, « Pains and Reasons: Why It Is Rational to Kill the Messenger », *The Philosophical Quarterly*, vol. 64, no. 256, 2014, p. 423-433.

———, « Tracking Representationalism and the Painfulness of Pain », *Philosophical Issues*, vol. 21, no. 1, 2011, p. 90-109.

Dennett, D. C., *Content and Consciousness*, International Library of Philosophy, London, Routledge, 1999.

Dennett, Daniel C., « Why You Can't Make a Computer That Feels Pain », *Synthese*, vol. 38, no. 3, 1978, p. 415-456.

Feldman, Fred, « Hyperventilating about Intrinsic Value », *Journal of Ethics*, vol. 2, no. 4, 1998, p. 339-354.

———, *Pleasure and the Good Life: Concerning the Nature, Varieties and Plausibility of Hedonism*, Oxford, Clarendon Press, 2004.

Frijda, Nico H., « On the Nature and Function of Pleasure », in Morten L. Kringelbach et Kent C. Berridge (dir.), *Pleasures of the Brain*, New York, Oxford University Press, 2009.

Goldstein, Irwin, « Hedonic Pluralism », *Philosophical Studies*, vol. 48, no. 1, 1985, p. 49-55.

———, « Why People Prefer Pleasure to Pain », *Philosophy*, vol. 55, no. 213, 1980, p. 349-362.

Hardcastle, Valerie Gray, *The Myth of Pain*, Cambridge, Mass., MIT Press, 1999.

Heathwood, Chris, « Desire Satisfactionism and Hedonism », *Philosophical Studies*, vol. 128, no. 3, 2006, p. 539-563.

———, « The Reduction of Sensory Pleasure to Desire », *Philosophical Studies*, vol. 133, no. 1, 2007, p. 23-44.

Jacobson, Hilla, « Killing the Messenger: Representationalism and the Painfulness of Pain », *The Philosophical Quarterly*, vol. 63, no. 252, 2013, p. 509-519.

———, « Not Only a Messenger: Towards an Attitudinal-Representational Theory of Pain », *Philosophy and Phenomenological Research*, vol. 99, no. 2, 2018, p. 382-408.

Kagan, Shelly, « The Limits of Well-Being », *Social Philosophy and Policy*, vol. 9, no. 2, 1992, p. 169-189.

Kammerer, François, « The Normative Challenge for Illusionist Views of Consciousness », *Ergo*, vol. 6, no.32, 2019, p.891-924.

Katz, Leonard D., « Pleasure », in Edward N. Zalta (dir.), *The Stanford Encyclopedia of Philosophy*, Winter 2016, Metaphysics Research Lab, Stanford University, 2016.

Katz, Leonard David, *Hedonism as Metaphysics of Mind and Value*, Dissertation, 1986.

Moore, G. E., *Principia Ethica*, traduit par Michel Gouverneur and Ruwen Ogien, Paris, PUF, 1998.

Nida-Rümelin, Martine, « Transparency and the Perceptual Model of Phenomenal Awareness », *Philosophical Perspectives*, vol. 21, no. 1, 2007, p. 429-455.

Parfit, Derek, *On What Matters*, The Berkeley Tanner Lectures, Oxford, Oxford University Press, 2011.

Pitcher, George, « Pain Perception », *Philosophical Review*, vol. 79, no. 3, 1970, p. 368.

Sidgwick, Henry, *The Methods of Ethics*, Indianapolis, Hackett Pub. Co, 1981.

Simon, Jonathan A., « Transparency About Painkillers: A Remedy for the Evaluativist's Headache », *Pacific Philosophical Quarterly*, vol. 100, no. 4, 2019, p. 935-951.

Smuts, Aaron, « The Feels Good Theory of Pleasure », *Philosophical Studies*, vol. 155, no. 2, 2011, p. 241-265.

Sumner, L. W., *Welfare, Happiness, and Ethics*, Oxford, Oxford University Press, 1996.

Tännsjö, Torbjörn, « Narrow Hedonism », *Journal of Happiness Studies*, vol. 8, no. 1, 2007, p. 79-98.

QUELLE PRO-ATTITUDE POUR LE SUBJECTIVISME À PROPOS DU BIEN-ÊTRE ?

STÉPHANE LEMAIRE

UNIVERSITÉ DE RENNES 1

RÉSUMÉ :

Selon le subjectivisme à propos du bien-être, un événement contribue au bien-être d'un individu dans la mesure où cet individu a ou aurait une certaine pro-attitude à l'égard de cet événement dans des circonstances à préciser. Une telle approche doit donc spécifier un certain nombre d'éléments de cette définition, par exemple les circonstances pertinentes. Dans cet article, je me concentre sur la pro-attitude pertinente. Afin d'organiser la discussion, je soutiens que cette pro-attitude doit satisfaire trois desiderata. Elle doit permettre de formuler une conception du bien-être qui serait extensionnellement adéquate moyennant une détermination satisfaisante des circonstances pertinentes. Elle doit offrir une conception du bien-être qui ne varie pas, peu importe que l'on considère des êtres humains à la naissance, jeunes ou adultes. Elle doit enfin exister effectivement. Ces desiderata posés, je montre qu'un subjectivisme faisant appel au désir, à une pro-attitude cognitive ou plus complexe, éventuellement comprise en termes de désirs, ne peut satisfaire à la fois le critère d'adéquation extensionnelle et le critère d'invariabilité. En réponse à cette difficulté, je suggère que le subjectivisme devrait adopter une pro-attitude de valorisation. Une telle pro-attitude a été introduite par différents auteurs, mais j'en propose une interprétation plus simple et je soutiens qu'elle est la seule à satisfaire les trois desiderata initialement présentés.

ABSTRACT:

Subjectivism as a view about well-being contends that an event contributes to an individual's well-being to the extent that this individual has or would have a certain pro-attitude towards this event under certain circumstances. To obtain a full-blown account of well-being, subjectivism must therefore specify several elements that are constitutive of the approach and, in particular, the relevant pro-attitude and circumstances. The present paper focuses on the relevant pro-attitude. It contends that this pro-attitude must satisfy three desiderata. It must provide a conception of well-being that could be extensionally adequate given a satisfactory determination of the relevant circumstances. It must offer a conception of well-being that applies equally well to infants, children, or adults. Finally, this relevant pro-attitude must exist. With these desiderata in hand, the paper shows that if subjectivism takes desire, or a cognitive attitude, or a more complex pro-attitude possibly understood in terms of desires, as relevant, then it cannot satisfy both the requirements of extensional adequacy and invariability. These difficulties suggest that subjectivism should look for a different pro-attitude. I propose after others that a pro-attitude of valuing is a better candidate. However, my account of this pro-attitude of valuing is simpler than those that have been offered so far and I argue that only this simpler understanding of valuing satisfies the three desiderata initially presented.

INTRODUCTION

Selon le subjectivisme à propos du bien-être, quelque chose contribue au bonheur ou au bien-être¹ d'un individu dans la mesure où ce dernier a une certaine attitude positive envers cette chose, ou en faveur de cette chose, dans certaines conditions. Historiquement, les auteurs subjectivistes ont considéré que l'attitude favorable (je dirai désormais pro-attitude) était le désir (Sidgwick, 1907 ; Brandt, 1998 ; Railton 1986a, 1986b). Ces auteurs ont par ailleurs soutenu que les conditions pertinentes n'étaient pas les circonstances actuelles des individus mais des circonstances idéales. Ainsi, Sidgwick a soutenu qu'un objet contribue au bien-être d'une personne, ou est bon pour cette personne, si et seulement si cette personne parfaitement informée sur la nature de cet objet désirerait cet objet (Sidgwick, 1907, p. 111). Dans ce qui suit, je ne m'intéresse pas à ces circonstances ; je ne me demande pas, en particulier, si elles doivent être idéalisées ou non. Je ne m'intéresse pas non plus à plusieurs autres questions qu'on peut se poser à propos du subjectivisme, par exemple : à quel moment la réalisation du contenu d'une pro-attitude contribue-t-elle au bien-être de la personne qui possède cette pro-attitude ? En effet, je souhaite m'intéresser exclusivement à cette autre question : quelle est la pro-attitude que devrait adopter le subjectivisme dans sa définition du bien-être ? Avant les années 1980, cette question était implicitement mise de côté dans la mesure où le désir se présentait comme l'unique candidat, mais depuis, plusieurs autres propositions ont émergé. On a d'abord suggéré que la pro-attitude pertinente puisse être une attitude complexe impliquant des désirs de premier et deuxième ordre (Railton, 1986a ; Lewis, 1989), puis le champ des solutions s'est encore élargi lorsqu'on a proposé que la pro-attitude pertinente puisse être strictement cognitive (Dorsey, 2012) ou encore une attitude de valorisation *sui generis* (Helm, 2001 ; Rosati, 2006 ; Tiberius, 2018). Dans ce dernier cas, l'idée est que ce qui contribue à notre bien-être n'est pas de réaliser simplement nos désirs, mais d'obtenir ce à quoi nous attachons une valeur.

Dans le présent article, je souhaite montrer qu'un certain nombre de considérations devraient pousser le subjectivisme à adopter la pro-attitude *sui generis* de valorisation. Cependant, l'analyse que j'en propose est plus simple que celles qui ont été proposées jusqu'ici. Selon moi, la bonne analyse de cette pro-attitude ne doit faire appel qu'à des dispositions émotionnelles, hédoniques et motivationnelles.

L'article se divise en quatre parties. Dans une première, je présente et justifie trois desiderata qui contraignent le choix de la pro-attitude pertinente pour le subjectivisme. Dans la seconde partie, je montre que le désir ne satisfait pas tous ces desiderata, et que les tentatives qui ont été faites pour contourner ces difficultés échouent. Dans la troisième partie, je me tourne vers la thèse selon laquelle une certaine attitude de valorisation serait celle que devrait adopter le subjecti-

visme. Je commence par présenter une conception simple de la valorisation et la distingue de versions plus sophistiquées. Je montre alors que seule cette conception simple de la valorisation permet au subjectivisme de satisfaire les trois desiderata mis en avant. Dans la quatrième partie, je réfute l'objection selon laquelle le subjectivisme ne pourrait pas utiliser la notion de valorisation sans renoncer à lui-même.

1. TROIS DESIDERATA

Comme je l'ai dit plus haut, la version initiale² du subjectivisme à propos du bien-être a été une forme de satisfactionnisme. Selon celui-ci, le bonheur consiste à réaliser ses désirs, tout au moins ceux qui sont ou seraient bien informés sur leur objet. La discussion de cette théorie et de toutes ses variantes a souvent porté sur leur adéquation extensionnelle. Ce sera donc notre premier desideratum. Par exemple, considérant le satisfactionnisme, on peut se demander s'il est nécessaire, pour qu'un objet soit bon pour celui qui le désire, que ce désir soi bien informé sur l'objet. On peut aussi se demander si la possession d'un désir bien informé est suffisante pour qu'un objet soit bon pour celui qui le désire. Bien sûr, un tel desideratum n'est que partiellement pertinent puisque nous ne cherchons pas ici à fournir une théorie subjective qui précise dans quelles circonstances le fait qu'un individu ait une certaine pro-attitude envers un objet en fait un objet qui contribue à son bien-être. Et en effet, tant que nous n'avons pas une théorie complète qui précise à la fois la pro-attitude et les circonstances pertinentes, la question de l'adéquation extensionnelle ne peut être pleinement discutée. Il n'en reste pas moins que l'adoption d'un certain type de pro-attitude pourrait conduire à une inadéquation extensionnelle qui ne pourrait être corrigée en modifiant les circonstances. Autrement dit, une pro-attitude, de par sa nature même, ne doit pas nous empêcher de reconnaître qu'un objet, une situation, un événement contribue au bien d'une personne, *quelle que soit la thèse que l'on pourrait adopter quant aux circonstances pertinentes*. Nous avons donc là un premier desideratum : la pro-attitude jugée pertinente ne doit pas rendre impossible l'adéquation extensionnelle du subjectivisme. On notera ici au passage que le concept de bien-être doit aussi permettre, conformément à l'usage ordinaire, de comparer des bien-être plus ou moins grands — voire des mal-être. Comme on le verra, le défaut de certaines théories satisfactionnistes est de ne pas pouvoir rendre compte de certaines variations dans le bien-être.

Le second desideratum qui sera pertinent est le desideratum d'invariabilité. En effet, on peut se demander si la notion de bien-être, de ce qui est bon pour un individu, désigne toujours la même propriété, qu'on s'intéresse aux êtres humains, aux animaux, aux végétaux ou même à des choses dépourvues de toute forme de vie. On ne s'intéressera pas ici à la thèse radicale — et d'ailleurs la plupart du temps rejetée — selon laquelle c'est toujours la même propriété à laquelle on fait référence.³ Ce que nous appellerons ici le desideratum d'inva-

riabilité est la thèse beaucoup plus modeste selon laquelle la même définition du bien-être s'applique à tous les êtres humains indépendamment de leur âge.⁴ Lin (2017, 2018) a bien montré la plausibilité de ce desideratum. On peut en saisir la portée en considérant un subjectivisme qui soutiendrait que le type de pro-attitude dont la satisfaction contribue au bien-être suppose des capacités cognitives complexes que ne possèdent pas les nourrissons. En effet, la définition même de ce subjectivisme implique soit que le bien-être pour les nourrissons qui ne possèdent pas ces capacités est différent, soit que les nourrissons n'ont pas de bien-être du tout — ce qui est une thèse absurde. Afin d'éviter cette dernière conclusion, ce subjectivisme doit donc rejeter la thèse de l'invariabilité : les pro-attitudes dont la réalisation contribue au bien-être ne sont pas du même type pour les adultes et les nourrissons. Or, comme l'a souligné Lin (2017), rejeter cette thèse a une conséquence inacceptable. En effet, elle implique que la satisfaction du type de pro-attitude qui contribue au bien-être du nourrisson ou du jeune enfant cesse complètement de contribuer à son bien-être dès qu'il acquiert les capacités plus sophistiquées qui impliquent qu'on lui applique la définition du bien-être des adultes. Or, il semble très difficile de comprendre ce qui pourrait expliquer une telle disparition. En effet, d'un point de vue métaphysique, si la satisfaction d'une certaine pro-attitude contribue au bien-être d'un individu, alors on ne comprend pas pourquoi la satisfaction de cette pro-attitude cesserait de contribuer au bien-être. Il semble qu'il faille alors soutenir que l'adulte ne possède plus ce type de pro-attitudes ou que celles-ci ne peuvent plus être satisfaites, ou enfin que les facultés sur lesquelles elles reposent sont détruites ou bloquées par l'apparition des capacités plus complexes. Cela est peut-être envisageable pour certaines pro-attitudes particulières, ou pour certaines facultés ou capacités, mais l'idée selon laquelle cela serait toujours le cas paraît très peu plausible, en particulier parce que le développement affectif et émotionnel de l'individu est progressif et n'implique pas de rupture totale et complète. Bref, il n'est pas plausible que, de toutes les pro-attitudes du type dont la satisfaction contribue au bien-être du nourrisson, aucune ne persiste ou ne puisse contribuer au bien-être de l'enfant plus âgé ou de l'adulte.⁵ Notre second desideratum sera donc un desideratum d'invariabilité : la pro-attitude que le subjectivisme doit adopter doit offrir une définition du bien-être invariable au sens au moins où elle s'applique à tous les êtres humains.

Le troisième et dernier desideratum qui va nous intéresser est un desideratum d'existence. En effet, une théorie subjectiviste du bonheur ne peut pas faire appel à une pro-attitude que les humains ne possèderaient pas. Historiquement, cette question ne s'est pas posée dans la mesure où l'existence de désirs n'a jamais été sérieusement remise en question si ce n'est aux jours les plus fougueux du behaviorisme. Mais si un subjectivisme fait appel à une pro-attitude de valorisation, alors il doit montrer que cette valorisation n'est pas qu'une notion séduisante, mais bien une pro-attitude qui existe réellement.

Ainsi, le présent article s'efforce d'évaluer dans quelle mesure un subjectivisme faisant appel à tel ou tel type de pro-attitude est en mesure de satisfaire les trois desiderata que nous venons d'exposer. Avant d'en venir à cette tâche, je voudrais toutefois souligner que je ne soutiens pas ici que ces trois desiderata sont les seuls que devrait satisfaire la pro-attitude pertinente pour le subjectivisme. En réalité, je pense qu'il existe au moins un autre desideratum d'ordre explicatif : l'appel à la pro-attitude pertinente devrait nous expliquer pourquoi son objet contribue au bien d'une personne. Malheureusement, montrer quelle pro-attitude permet effectivement de réaliser ce desideratum nécessite une discussion qui ne peut être conduite dans le présent article.⁶

2. LE SUBJECTIVISME EN TERMES DE DÉSIRS ET AU-DELÀ

L'ambition de cette section est de montrer qu'un certain nombre de théories subjectives faisant appel à une pro-attitude conative plus ou moins sophistiquée ou cognitive n'arrivent pas à satisfaire les trois desiderata que nous avons mis en avant. Je m'intéresse d'abord à la théorie la plus simple en termes de désirs de premier ordre, puis je me tourne vers des théories plus complexes avant de considérer une approche strictement cognitive.

2.1 Comprendre le bien-être en termes de désirs satisfaits

Selon le satisfactionnisme, ce qui est bon pour une personne est ce qu'elle désire dans sa situation actuelle, ou dans des conditions idéales.⁷ Conformément à l'orientation du présent article, je ne vais pas m'intéresser au débat entre les défenseurs d'une position actualiste et ceux qui au contraire militent en faveur de l'introduction d'une condition d'idéalisation. Je me concentre seulement sur la possibilité pour un satisfactionnisme en termes de désirs de satisfaire les trois desiderata que nous avons présentés.

Concernant le desideratum d'existence, on voit immédiatement qu'il est satisfait puisque, comme je l'ai déjà dit, l'existence des désirs n'est pas remise en question. Qu'en est-il alors du desideratum d'invariabilité. Est-il satisfait ? Il ne semble y avoir aucune raison de supposer qu'il ne l'est pas. En effet, s'il est admis que les choses qui sont bonnes pour une personne sont celles qui sont désirées par cette personne, alors le fait que les nourrissons comme les adultes possèdent des désirs permet de comprendre leur bonheur de la même manière et en continuité.⁸

Restent les objections concernant l'adéquation extensionnelle. Une première liste d'objections vise à montrer que la possession d'un désir envers un objet n'est pas une condition suffisante pour que cet objet contribue au bien-être de la personne qui possède ce désir. Par exemple, on a pu objecter que la réalisation de désirs que l'on éprouve comme des besoins irrépressibles mais vides de sens

ne contribue pas au bien-être. Le défenseur d'un subjectivisme faisant appel aux désirs peut toutefois souvent répondre en précisant quels sont les désirs pertinents ou en introduisant une dose d'idéalisation. Il peut répondre avec Heathwood (2019) que les désirs pertinents sont seulement des désirs attractifs, c'est-à-dire ceux qui impliquent que l'on éprouve une attraction pour l'objet en question soit du fait d'éprouver des émotions relativement à la satisfaction de ce désir, soit d'envisager la réalisation de ces désirs comme nous apportant du plaisir. Une autre option sera de répondre que seuls les désirs que nous assumons ou auxquels nous nous identifions (Frankfurt, 1971 ; Bruckner, 2016) déterminent le bien-être. Enfin, une troisième réponse est de soutenir que ces désirs n'existent pas réellement et ne sont que des fictions philosophiques. Après tout, s'il est dur de résister à de telles impulsions, c'est parce que cette résistance a un certain coût qui les rend négativement attractifs au sens de Heathwood. Bref, on se débarrasserait du problème en disant tout simplement que les désirs qui constituent des contre-exemples en réalité n'existent pas et ne sont que des fictions philosophiques.

On a pu soutenir aussi que des désirs sadiques par exemple ne peuvent pas contribuer au bien-être. Mais à nouveau, le subjectiviste a de nombreuses options pour répondre. Il peut répondre soit que, bien que moralement mauvais, la satisfaction de ces désirs reste *pro tanto* bonne pour celui qui possède ces désirs (Heathwood, 2005), soit soutenir qu'il faut ajouter au moins pour certains désirs une condition objective (Wall et Sobel, 2021).

Bref, il ne semble pas que l'idée selon laquelle le désir serait la pro-attitude pertinente soit vraiment mise en danger par ces prétendus contre-exemples. La menace sur l'adéquation extensionnelle est par contre plus sérieuse si l'on se tourne vers des cas où la possession d'un désir à l'égard d'un événement ne semble pas nécessaire pour que ce dernier contribue au bien-être. Heathwood considère l'exemple suivant. Sa femme s'approche de lui et, sans qu'il l'anticipe, lui masse les épaules. Il éprouve alors un plaisir immédiat, une merveilleuse détente, laquelle contribue sûrement à son bien-être à ce moment-là. Pour autant, avoue-t-il, il n'avait pas le désir qu'elle lui masse les épaules avant qu'elle ne le fasse. Plus généralement, il semble indiscutable que de nombreuses activités, sensations, nous apportent du plaisir par surprise pour ainsi dire, et donc sans que nous ayons, au moins explicitement, désiré faire ces activités, ressentir ces sensations.

Afin d'éviter ces contre-exemples, Heathwood (2006, p. 558) propose la réponse suivante : celle-ci consiste à reconnaître en premier lieu qu'il n'avait certes pas de désir d'être massé avant de l'être. Néanmoins, un désir d'être massé s'est formé au moment même où il est massé. Dès lors, on peut expliquer selon le schéma subjectiviste standard la contribution au bien-être qui résulte du massage puisque, au moment où Heathwood est massé, il possède le désir d'être massé et il est massé.

Cette réponse me paraît toutefois insatisfaisante pour deux raisons. En premier lieu, si l'expérience initiale et agréable du massage engendre dès que je l'éprouve le désir que le massage se poursuive, il n'est pas plausible qu'à l'instant initial où je suis massé et en éprouve du plaisir, je possède un désir d'être massé à ce même instant précis. S'il est plausible que nous acquérions immédiatement un désir, le désir en question est bien plutôt que ce massage *dure* encore. Autrement dit, le désir qui apparaît au moment où le massage commence est un désir qui porte sur le futur, même s'il s'agit du futur le plus immédiat. De façon générale, s'il est plausible de dire que lorsque j'ai un désir, celui-ci perdure au moins une partie du temps de sa réalisation de telle sorte que pendant une certaine période je continue de posséder un désir alors qu'il est en train de se réaliser — considérez par exemple le désir de manger, ou de courir —, il est un peu curieux de dire que pourrait naître un désir qui porterait spécifiquement sur ce qui advient à l'instant présent au sens strict. En effet, dans la mesure où de tels désirs seraient toujours par définition satisfaits, on se demande bien à quoi ils serviraient si ce n'est à sauver la théorie de Heathwood. Pourquoi la nature nous aurait-elle dotés de tels désirs ?

En second lieu, il y a quelque chose d'insatisfaisant dans la réponse de Heathwood. En effet, le propre du subjectivisme est de présenter une pro-attitude relativement stable et de soutenir que lorsque l'objet de la pro-attitude est réalisé ou lorsque le sujet croit que cet objet est réalisé, alors cela contribue au bien-être de la personne qui possède cette pro-attitude. Mais la stratégie employée par Heathwood pour sauver sa théorie des cas de plaisir sans désir antérieur renonce à cette dimension explicative. Pour conserver la vertu explicative du subjectivisme, il faudrait expliquer à la fois le plaisir qui est éprouvé et le désir qui apparaît au même moment. Or, pour cela, il semble qu'il faille faire appel à un autre élément qui serait quelque chose comme une disposition à avoir ensemble le plaisir et le désir que celui-ci perdure. Mais si on fait appel à un tel état mental pour expliquer la co-occurrence du plaisir et du désir, ne faudrait-il pas suggérer que celui-ci constitue la pro-attitude que le subjectiviste devrait adopter ? En outre, si l'objection précédente est correcte, alors la seule pro-attitude dont la réalisation va de pair avec ces plaisirs non anticipés est précisément cette disposition sous-jacente, et non le désir. Bref, non seulement la stratégie de Heathwood échoue, mais, en outre, elle fait perdre au subjectivisme sa dimension explicative. Enfin, elle semble indiquer une autre pro-attitude que le subjectivisme pourrait adopter, qui lui éviterait le problème des plaisirs non anticipés et lui conserverait sa dimension explicative.

Cependant, une réponse plus prometteuse que celle d'Heathwood ne serait-elle pas d'avancer que nous avons le désir général d'avoir du plaisir, et donc que le massage étant plaisant, il satisfait ce désir ?⁹ Mais avons-nous un tel désir général ? Pour le voir, on peut commencer par évaluer le cas partiellement symétrique de la douleur. Avons-nous le désir général d'éviter la douleur ? Est-ce que

chaque fois que nous avons une expérience de douleur, nous cherchons à l'éviter en raison de ce désir général antérieur ? Cela semble extrêmement peu plausible et apparaît plutôt comme une façon inadéquate de décrire nos réactions à la douleur. Bien sûr, chaque fois que j'éprouve de la douleur, j'ai le désir que cette douleur cesse. Mais cela n'est pas parce que j'ai le désir général d'éviter la douleur, mais simplement parce que la douleur est ce type d'expérience qui engendre ou implique constitutivement le désir que cette douleur que j'éprouve cesse. Bien sûr, le fait d'éprouver une certaine douleur par exemple au genou peut engendrer un désir durable de ne pas vivre à nouveau cette douleur au genou et par conséquent de faire ce qu'il faut pour l'éviter. Mais toutes les douleurs ne supposent pas un tel désir, en particulier celles auxquelles je ne m'attends pas. En outre, je peux rechercher certaines douleurs, par masochisme ou parce qu'elles sont l'occasion d'épreuves auxquelles je souhaite me soumettre, ou que je souhaite traverser. Dans ce cas, il est loin d'être clair que j'ai aussi le désir antérieur de ne pas les éprouver.

Ces remarques se transposent au plaisir. Un plaisir engendre sans doute constitutivement un désir, mais la plupart d'entre nous ne désire pas le plaisir en général. En particulier, certains ne désirent pas éprouver certains plaisirs, parce qu'ils les jugent dégradants. Il n'est donc pas vrai que nous désirions le plaisir de façon générale. Et si certains peuvent dire qu'ils désirent tous les plaisirs, alors c'est une éthique de vie, un choix qui n'est pas nécessaire, et qui donc ne permet pas de conclure en général et pour tout être humain que le plaisir est toujours désiré par avance. Je conclus que cette réponse ne s'avère en réalité pas plus prometteuse, et ainsi que l'existence de plaisirs non anticipés pose un problème d'adéquation extensionnelle au satisfactionnisme, tout au moins tel que nous l'avons compris jusqu'ici.

2.2 Un satisfactionnisme plus sophistiqué ?

Pourrait-on sauver le satisfactionnisme en qualifiant les désirs qui contribueraient au bien-être ? Selon une suggestion qui est souvent considérée comme une amélioration substantielle du satisfactionnisme, un objet est bon pour un individu s'il est l'objet d'un désir *assumé* par cet individu dans certaines conditions, ou s'il possède un désir de second ordre à l'égard de son désir de premier ordre.¹⁰ Comme on l'a vu plus haut, un avantage immédiat de cette suggestion est que l'on peut écarter les désirs irrépressibles mais dont la satisfaction nous est intuitivement indifférente, car, du fait même de concevoir leur satisfaction comme indifférente, voire étrangère, ces désirs ne seront pas assumés et ne seront pas l'objet d'un désir de second ordre.

Il y a toutefois deux raisons de rejeter ces satisfactionnismes sophistiqués. En premier lieu, ces versions du satisfactionnisme ne peuvent résoudre le problème posé par les plaisirs non anticipés puisqu'on ne peut pas assumer un désir qu'on

ne possède pas. En outre, l'introduction de la condition d'assomption conduit à un autre problème d'adéquation extensionnelle important. Considérons l'exemple d'une personne homosexuelle, qui vit dans un environnement rejetant violemment l'homosexualité et qui, de ce fait, est amenée elle-même à rejeter violemment sa propre homosexualité. Dans ce cas, elle n'assume pas ses désirs homosexuels et il faudrait conclure que ni la satisfaction ni la frustration de ses désirs sexuels ne contribue, même *pro tanto*, à son bien-être. Mais cette conclusion est évidemment fautive. Dans les circonstances qui sont les siennes, il ne fait pas de doute que la satisfaction de son désir homosexuel va contribuer à son bien-être ; en réalité elle y contribue de deux manières : premièrement, de façon *pro tanto* positive ; deuxièmement, du fait de ne pas assumer son désir tout en l'ayant — et *a fortiori* en le satisfaisant — de façon *pro tanto* négative. Notons en outre que ces contributions *pro tanto* s'annuleront rarement si jamais cela advient. Bref, le bien-être de cette personne qui est homosexuelle tout en le rejetant et qui satisferait son désir homosexuel n'aura rien à voir avec celui d'une personne qui ne posséderait pas de désir homosexuel du tout.

Est-ce que ce problème pourrait être évité si on supposait que les désirs dont la satisfaction contribue au bien-être sont ceux qu'on assumerait à l'issue d'une réflexion idéale ? Dans l'exemple précédent, on pourrait supposer que le désir homosexuel serait alors assumé. Dans ce cas, sa réalisation contribuerait au bien-être. Mais par ailleurs, comme l'évaluation idéalisée aurait pour effet d'éliminer le désir de second ordre de ne pas avoir de désir homosexuel lié à son refus d'assumer son désir homosexuel, la possession de son désir homosexuel ne nuirait nullement à son bien-être. Or, cela est à nouveau incorrect. Si nous rejetons fortement un de nos désirs, alors sa simple existence nuit indiscutablement à notre bien-être, à notre paix avec nous-mêmes. Même si la satisfaction de notre désir de premier ordre contribue *pro tanto* à notre bien-être, l'évaluation de second ordre et le désir corrélé à cette évaluation continueront d'engendrer des émotions ou sentiments mêlés qui seront la marque d'une contribution négative à notre bien-être.

Pour finir, comme l'a montré Lin (2017, p. 361), le subjectivisme sophistiqué ne peut pas satisfaire le desideratum d'invariabilité. En effet, sa définition même implique que pour qu'un objet contribue au bien d'une personne, il est nécessaire que celle-ci possède des capacités cognitives lui permettant d'évaluer ses propres désirs ou qu'elle possède des désirs de second ordre. Or, les nourrissons ne possèdent pas de telles capacités cognitives ni non plus de désirs de second ordre. Dès lors, ils n'auraient pas de bien-être. Pour éviter cette conclusion absurde, le défenseur doit donc rejeter la thèse d'invariabilité. Mais nous avons montré que cette hypothèse n'est pas défendable.

En résumé, un satisfactionnisme plus sophistiqué introduisant une évaluation cognitive ou des désirs de second ordre, une forme d'approbation ou d'assomp-

tion de nos désirs ne nous fait pas progresser face aux problèmes des plaisirs et émotions non anticipés, et s'il résout le problème des désirs irrésistibles, ce n'est que pour faire apparaître un problème symétrique pour des désirs qu'on n'assumerait pas. En outre, ces satisfactionnismes sophistiqués ne satisfont pas le desideratum d'invariabilité.

2.3 Sauver le satisfactionnisme par l'idéalisation

Nous venons de voir qu'un satisfactionnisme qui introduit une évaluation des désirs ou des désirs de second ordre satisfait en réalité moins bien les desiderata d'adéquation extensionnelle et d'invariabilité que le satisfactionnisme actualiste tel qu'il est défendu par Heathwood par exemple. Pourrait-on envisager une autre façon de sauver le satisfactionnisme en introduisant des clauses idéalisantes ? On peut envisager deux types de clauses idéalisantes. Selon un premier type, les désirs dont la satisfaction est pertinente pour le bien-être sont les désirs qui sont centraux, ou ceux qui d'une façon ou d'une autre sont cohérents entre eux ou seraient cohérents à l'issue d'une procédure de mise en cohérence. Cependant, ce premier type de stratégie ne règle pas le problème des plaisirs non anticipés puisqu'il n'aura pas pour effet de faire émerger de façon systématique des désirs portant sur tous les plaisirs que nous n'anticipons pas. Selon le second type d'idéalisation, les désirs dont la satisfaction contribuerait au bien-être sont ceux qu'une personne aurait si elle était parfaitement informée et/ou si elle avait fait l'expérience de cette chose. L'avantage d'une telle clause idéalisante est qu'elle permet de résoudre le problème des plaisirs non anticipés. En effet, si Heathwood avait fait l'expérience du massage qu'il a reçu, alors il est plausible qu'il en aurait par avance le désir. Le problème serait résolu.

L'appel à une information idéale et/ou à une expérience maximale pose toutefois des problèmes importants qui ont été soulignés dans la littérature. Le premier, et le plus évident, est qu'on ne sait pas très bien quels seraient véritablement les désirs d'une personne ainsi informée (Gibbard, 1990 ; Arneson, 1999). Il se pourrait que l'appel à une information idéale ne débouche pas sur le résultat escompté. Peut-être que Heathwood ayant fait l'expérience de nombreux plaisirs sensuels serait assez déçu par le massage de son épouse. Ce point souligne que les plaisirs et émotions en particulier vis-à-vis de la nouveauté sont sensibles à l'ordre de présentation, et que la contribution d'un objet à notre vie va dépendre de sa relative nouveauté. Or, ce point est totalement perdu en faisant appel à une expérience idéalisée ou complète (Rosati, 1995 ; Sobel, 1994). Cet écart sera si considérable si on s'intéresse à ce qui contribuerait au bien-être d'un nourrisson qu'on se demande bien en quoi la réalisation de désirs qu'il aurait à l'issue d'une longue expérience et d'une pleine information pourrait effectivement affecter le nourrisson lui-même (Lin 2017, p. 365-367). En somme, si le satisfactionnisme introduit une idéalisation informationnelle forte, alors il ne satisfera pas la condition qu'on appelle parfois de résonance qui est

si centrale à la défense du subjectivisme et selon laquelle un objet ne peut contribuer au bien-être d'un individu que si cet objet est susceptible d'affecter l'individu d'une manière ou d'une autre. Autrement dit, un satisfactionnisme avec une clause d'idéalisation informationnelle forte ne satisfait pas la condition de résonance et de ce fait la condition d'adéquation extensionnelle.¹¹ Mais, à l'inverse, si la clause informationnelle est faible et vise simplement à identifier les objets qui contribueront effectivement à notre bien-être (Sobel, 2009), alors le problème des plaisirs non anticipés montre que la pro-attitude pertinente ne saurait être le désir.

Afin d'éviter ces difficultés, une variante de cette dernière idée serait de voir l'individu idéalement informé non pas comme un modèle, mais comme un conseiller suggérant à la personne actuelle ce qui serait bon pour elle. On soutiendrait alors que les objets qui contribuent au bien-être d'une personne sont ceux que la personne qui serait omnisciente et rationnellement pratique conseillerait à la personne actuelle de désirer (Smith, 1995) ou désirerait que cette personne désire (Railton, 1986a). L'avantage de cette position est que la personne à propos de qui on évalue ce qui est bon pour elle n'est pas transformée par des expériences ou informations qu'on lui donne.

Cette dernière approche a toutefois le défaut d'être largement indéterminée. On ne sait pas ce qu'une personne idéalement informée conseillerait à son moi actuel tant qu'on n'a pas clarifié ce qui guiderait ses conseils à son moi actuel, ou désirerait que son moi actuel désire. En réalité même, la réponse la plus plausible est qu'elle conseillerait de réaliser les choses qui contribueront à leur bien-être. Bref, si l'idée d'un conseil fournit des analyses intéressantes, celles-ci restent circulaires. Elles ne déterminent pas ce qui ontologiquement contribuera au bien-être d'un individu du point de vue subjectiviste, mais servent au mieux d'aide du point de vue épistémique pour évaluer des réponses à la question ontologique (Rosati, 2006, p. 110).¹²

Bref, si l'avis d'un individu idéalement informé est pertinent, c'est d'abord, semble-t-il, parce que celui-ci est susceptible de savoir quelles sont les choses qui contribueront effectivement au bien-être d'une personne étant donné ses dispositions subjectives. Ainsi, cette personne saura — ce qu'Heathwood lui-même ne sait pas — qu'un massage à ce moment lui fera du bien. Mais comme nous l'avons vu, si l'état nous dispose à tirer du bien-être d'un massage n'est pas le désir de ce massage mais une autre pro-attitude, alors il faut conclure que derrière l'appel indirect à des conseils idéaux, le désir n'est pas ontologiquement ce qui constitue le bien-être. En d'autres termes, l'appel à un conseiller idéal ne ferait rien d'autre que cacher le fait qu'en fin de compte un satisfactionnisme ne peut pas être extensionnellement adéquat, notre premier desideratum.

2.3 Un subjectivisme cognitif ?

Si on doit rejeter les différentes versions de la thèse selon laquelle la pro-attitude pertinente pour le subjectivisme est le désir, alors ne peut-on envisager que cette pro-attitude soit plutôt d'ordre cognitif ? Pourquoi pas ? Cette suggestion suppose néanmoins qu'on identifie l'attitude cognitive que le subjectivisme pourrait adopter. La suggestion sans doute la plus plausible — et de fait la seule qui a été suggérée — est que cette pro-attitude soit le jugement ou la croyance qu'une certaine chose est bonne pour nous (Dorsey, 2017). En effet, si nous croyons qu'une chose est bonne pour nous, alors il est probable qu'elle le soit effectivement.

Un avantage de cette position est qu'elle évite certains des contre-exemples présentés plus haut. Par exemple, si nous avons une pulsion qui nous pousse à faire quelque chose sans que nous y trouvions le moindre sens, alors nous jugerons sans doute que sa réalisation n'est pas bonne pour nous. On peut ainsi se débarrasser d'un certain nombre de cas qui constituaient des contre-exemples au satisfactionnisme.

En outre, et dans le sens inverse, cette approche cognitiviste évite l'objection des plaisirs non anticipés. Même si je ne désire pas avoir un plaisir avant d'en faire l'expérience, je forme sûrement, ou suis disposé à former, au moment où je fais une expérience plaisante, le jugement que cette expérience est bonne pour moi.

L'approche cognitiviste ne satisfait malheureusement pas les desiderata d'invariabilité et d'adéquation extensionnelle pour des raisons que nous avons déjà vues. Comme Lin (2017) l'a souligné, le desiderata d'invariabilité est violé par le cognitivisme de Dorsey parce que cette définition du bien-être ne peut s'appliquer aux êtres qui n'ont pas les capacités cognitives pour former des jugements sur ce qui est bon pour eux. Dès lors, Dorsey doit soit admettre que les nouveau-nés n'ont pas de bien-être du tout, ce qui semble absurde, soit admettre que pour ces êtres une conception différente du bien-être doit s'appliquer et donc que le desiderata d'invariabilité n'est pas satisfait. En réponse à cette objection, Dorsey (2017) a soutenu que ce desideratum n'existe que pour autant qu'on ne peut pas justifier la variation de ce qui est désigné par ce qui est bon pour un individu et Dorsey s'efforce de justifier une telle variation. Cette stratégie est toutefois confrontée à une difficulté majeure. En effet, supposons qu'on l'adopte et considérons des objets qui contribuent au bien-être d'un individu avant et après que soit changée la définition du bien-être qu'il est pertinent de lui appliquer. On peut prendre pour exemple la satisfaction des besoins alimentaires ou affectifs. Dans de tels cas, l'explication de leur contribution au bien-être sera différente selon le moment auquel on se place. Mais, cela semble très bizarre. Pourquoi l'apparition de nouvelles facultés annule le rôle de celles qui expliquaient le bien-être avant leur apparition ?

En second lieu, la théorie cognitive est confrontée au même problème d'inadéquation extensionnel que les versions sophistiquées du satisfactionnisme. En effet, il est tout à fait clair que l'homosexuelle malgré elle jugera qu'avoir des relations homosexuelles n'est pas une bonne chose pour elle. Pourtant, il est assez indiscutable que ce serait au contraire une chose à la fois *pro tanto* bonne et *pro tanto* mauvaise pour elle si elle satisfaisait ce désir comme on l'a vu plus haut.¹³

Bien sûr, si aucune théorie subjectiviste ne pouvait éviter cette difficulté ou si son évitement conduisait toujours à d'autres difficultés, cela pourrait justifier qu'on adopte le cognitivisme ou une des approches qu'on a vues plus haut. Je voudrais toutefois essayer de montrer maintenant qu'un subjectivisme en termes de valorisation est capable de mieux satisfaire les trois desiderata que nous avons présentés.

3. VALORISER

Dans cette section, j'introduis la pro-attitude qui me paraît la plus à même de rendre plausible une approche subjectiviste du bonheur. Je propose que cette pro-attitude soit celle consistant à valoriser une chose, un événement, une situation, etc. Une pro-attitude de valorisation a été introduite dernièrement par différents auteurs avec des caractérisations différentes et avec des objectifs théoriques différents. Dans ce qui suit, je définis tout d'abord ce que j'appelle la conception simple de la valorisation au moyen d'un certain nombre de traits. Je montre ensuite que cette conception simple de la valorisation offre une forme de subjectivisme qui satisfait les trois desiderata que j'ai introduits alors que les subjectivismes qui s'appuient sur des conceptions plus sophistiquées de la valorisation échouent à satisfaire ces desiderata.

3.1 Valoriser : la conception simple et le desideratum d'existence

Je propose de concevoir la pro-attitude de valorisation de la façon suivante :

Conception simple de la valorisation :

Un individu valorise un objet (événement, état de choses, etc.) si et seulement s'il est disposé à éprouver des émotions ou du (dé)plaisir vis-à-vis de cet objet et s'il est disposé à avoir une motivation relativement à cet objet.

Cette conception constitue à mon sens la bonne façon de caractériser la valorisation, en tous les cas, lorsqu'il s'agit d'identifier la pro-attitude pertinente pour une approche subjectiviste du bien-être. L'attitude de valorisation en ce sens existe-t-elle bel et bien ? À mon sens, il existe au moins deux raisons en faveur de l'existence de valorisations conçues simplement.

En premier lieu, il faut noter que cette conception est implicite lorsqu'on dit ordinairement que l'on valorise quelque chose. Par exemple, si je dis que je valorise mes relations familiales, alors cela veut dire que je suis disposé à cultiver ces relations, que j'en tire du plaisir, des émotions positives, mais que je suis aussi inquiet si ces relations se dégradent d'une manière ou d'une autre, et que je suis corrélativement disposé à faire un certain nombre de choses pour qu'elles ne se dégradent pas et pour qu'elles perdurent. À l'inverse, en l'absence de ces dispositions, on dira certainement que je ne fais que prétendre les valoriser et que ce ne sont en réalité pas des valeurs pour moi. Ceci dit, ce à quoi on réfère ordinairement en parlant de ce qu'on valorise implique des traits qui vont au-delà de la conception simple. En particulier, un sujet qui valorise quelque chose va admettre qu'il assume ses dispositions motivationnelles et émotionnelles relativement à un objet ou reconnaît que l'objet de cette valorisation est bon pour lui ou bon *simpliciter*. Quoi qu'il en soit, et quelle que soit la suggestion que l'on retienne, ce qui nous importe ici est que la valorisation ainsi comprise implique toujours des dispositions émotionnelles et motivationnelles vis-à-vis d'un objet. Autrement dit, l'usage ordinaire de la notion de valorisation fait appel implicitement à des valorisations conçues simplement puisque c'est à partir de ces dernières, une fois reconnues, que nous identifions ce qui est, au sens ordinaire, valorisé. Bref, même si la conception simple de la valorisation est dans une certaine mesure révisionniste, la notion ordinaire reconnaît implicitement son existence.

Une seconde raison d'admettre l'existence de valorisations (comprises selon la conception simple) repose sur l'explication de la forte corrélation que l'on trouve entre les émotions et les désirs. D'un côté, nous savons qu'une émotion implique toujours une motivation, soit que l'on considère que les émotions sont des états motivationnels (Frijda, 2007 ; Scarantino, 2014) soit qu'on considère que les motivations sont une composante des émotions,¹⁴ soit que l'on considère enfin qu'il existe une relation causale étroite (Tappolet, 2016). À l'inverse, la possession d'un désir implique des dispositions émotionnelles relativement à la satisfaction ou la frustration de ce désir ou à leur anticipation. Comment expliquer ces corrélations étroites ? Une option serait de considérer que l'un de ces états est premier et nous dispose à être dans l'autre type d'état. Mais il est difficile de dire que les désirs sont plus fondamentaux dans la mesure où ils apparaissent souvent du fait d'une émotion. En outre, les cas de plaisirs et émotions non anticipés nous ont montré qu'il n'y avait pas alors de désir antérieur. Les émotions sont-elles plus fondamentales ? Cela semble tout aussi difficile à défendre dans la mesure où elles supposent au moins parfois un désir pour les expliquer. Par ailleurs, les théories des psychologues visant à expliquer nos émotions admettent pour la plupart que les émotions impliquent ou résultent de l'évaluation de circonstances relativement à ce qu'on pourrait appeler de façon générale des buts ou intérêts (Ellsworth et Scherer, 2001 ; Ortony, Clore et Collins, 1990 ; Frijda, 2007). Par exemple, la peur résulterait ou impliquerait systématiquement

l'évaluation d'un objet comme une menace sur ce qui nous importe. Notez en outre que ces buts ne sont pas nécessairement des désirs puisque nous avons des émotions et plaisirs sans désir antérieur. Dès lors, il semble que la suggestion suivante se présente : si les buts postulés par les théoriciens des émotions expliquent nos dispositions émotionnelles, lesquelles impliquent aussi la co-occurrence de motivations, pourquoi ne pas identifier ces buts à des valorisations et les concevoir comme étant à la fois des dispositions émotionnelles et des dispositions à former des désirs ?

On pourrait s'opposer à cette thèse si nous pouvions avoir des désirs indépendamment de nos dispositions émotionnelles ou l'inverse. En effet, on pourrait alors plausiblement distinguer d'un côté nos dispositions émotionnelles et de l'autre nos dispositions à former des désirs. Mais de tels cas ne semblent pas exister. La formation de nouveaux désirs est liée systématiquement à des expériences émotionnelles. Ce par quoi je m'attribue un désir est souvent une émotion. Parfois cette émotion révèle le désir, mais en d'autres cas, le désir apparaît avec l'émotion. Si j'ai un désir, j'ai des dispositions émotionnelles, et si j'ai des dispositions émotionnelles, alors j'ai la disposition à avoir des désirs corrélés à ces émotions. Dès lors, il faut reconnaître que l'apparition de nouveaux désirs repose sur les mêmes mécanismes que ceux qui sont responsables de nos émotions et donc ultimement sur l'existence de buts ou d'intérêts. On notera d'ailleurs que parce que les désirs impliquent eux-mêmes des dispositions émotionnelles, ils doivent compter de ce fait parmi les intérêts qui expliquent nos émotions. Après tout, nous avons peur que notre désir ne se réalise pas, nous nous réjouissons qu'il se réalise, etc.

Il est donc plausible que nos émotions et motivations procèdent les unes et les autres des mêmes états mentaux sous-jacents, lesquels états mentaux se caractériseraient précisément du point de vue fonctionnel comme des dispositions à avoir à la fois des émotions et des désirs relativement à certains objets intentionnels.¹⁵ C'est ainsi d'ailleurs que j'ai proposé de concevoir la valorisation.

Afin de clarifier cette conception simple de la valorisation, je voudrais pour finir m'attarder sur une objection qu'on pourrait lui faire. En effet, on pourrait être tenté de rejeter l'idée selon laquelle toutes les valorisations impliquent une disposition motivationnelle comme je le prétends. Après tout, il est tout à fait courant de valoriser ce qu'on a accompli dans le passé. Or, remarque plausiblement l'objecteur, on n'a pas de motivation vis-à-vis de ces accomplissements, et pas non plus de disposition à avoir de telles motivations.¹⁶

En réponse, il faut certainement concéder que nous n'acquérons pas de motivations vis-à-vis d'un passé valorisé car, dans la mesure où celui-ci, par nature, n'est pas modifiable, nous n'avons pas à nous inquiéter comme pour d'autres choses que nous valorisons de sa possible disparition. S'ensuit-il pour autant

que nous ne sommes pas disposés à désirer que ce passé ne change pas s'il était possible qu'il change ? Il y a plusieurs raisons de répondre négativement. En premier lieu, même s'il est métaphysiquement impossible de changer le passé, nous pourrions être amenés à croire que c'est possible. Supposons ainsi que nous soyons amenés à penser qu'un certain passé valorisé pourrait être modifié de telle façon que nous n'aurions plus fait ce que nous valorisons tout en continuant de le valoriser. Il semble que dans ce cas nous désirerions que ce passé ne change pas. Autrement dit, contrairement à ce que prétend l'objection, nous avons des dispositions motivationnelles vis-à-vis de notre passé, même si elles n'ont de fait aucune raison de s'actualiser. Ce point se trouve renforcé par deux considérations supplémentaires. Tout d'abord, on peut considérer le cas symétrique dans lequel nous valorisons négativement, c'est-à-dire dévalorisons, ce que nous avons fait. Dans ce cas, il ne fait pas de doute que nous désirons ne pas avoir fait ce que nous avons fait. Par ailleurs, il est tout à fait possible de valoriser des actions que j'aurais pu faire dans le passé mais que je n'ai pas faites. Par exemple, je peux valoriser le fait d'avoir combattu dans la résistance et regretter de ne pas l'avoir fait. Or, regretter de ne pas l'avoir fait, c'est désirer ou être disposé à désirer l'avoir fait. Bref, le fait de valoriser un certain passé implique un désir, soit parce que cette valorisation est négative et que j'ai réalisé ce passé, soit parce que la valorisation est positive alors que je n'ai pas réalisé ce passé. Il en résulte plausiblement par parité que la valorisation positive d'un passé réalisé implique aussi une disposition à agir même si celle-ci n'a aucune raison de s'actualiser.

Supposons donc que la valorisation conçue simplement satisfasse le desideratum d'existence, la question qu'on peut dès lors se poser est celle-ci : une approche subjectiviste du bien-être devrait-elle faire appel à cette conception simple ou à une conception plus sophistiquée et par là même peut-être plus proche du sens commun ? Dans ce qui suit, je m'efforce de montrer que la conception simple de la valorisation l'emporte sur les conceptions plus sophistiquées qu'on pourrait en proposer au regard des desiderata d'adéquation extensionnelle et d'invariabilité.

3.2 Adéquation extensionnelle

En premier lieu, je voudrais souligner qu'un subjectivisme faisant appel à une conception simple de la valorisation est préférable à l'approche satisfactionniste parce qu'il est à même de répondre au problème des plaisirs non anticipés auquel le satisfactionniste ne pouvait répondre. En effet, même si certains désirs sont des valorisations ou impliquent de façon sous-jacente une valorisation, valoriser quelque chose n'implique pas nécessairement qu'on désire effectivement cette chose puisque le propre de la valorisation est d'être une disposition à désirer qui peut être réalisée ou ne pas l'être. Ainsi, l'appel à la valorisation permet de reconnaître que des plaisirs non anticipés, non désirés, contribuent à notre bien-être.

Ils y contribuent parce que nous les valorisons. Nous avons de façon innée une disposition à apprécier le fait d'être touché, massé ou caressé, même si notre histoire nous conduit certainement à préférer l'être d'une façon plutôt que d'une autre et dans un contexte plutôt que dans un autre. De la même manière, les plaisirs esthétiques sont d'abord inattendus. Je peux écouter une musique que je ne connais pas — par exemple un raga indien — et m'apercevoir que je l'apprécie, parce que cette écoute me retient et m'est plaisante. On peut toutefois reconnaître qu'à ce moment, j'avais la ou les dispositions me permettant d'apprécier cette musique. Bref, je valorisais, au sens où j'emploie le terme, cette musique. Certes, je ne savais pas avant d'écouter cette musique que j'avais cette disposition et, bien sûr, du fait d'éprouver du plaisir en l'écoutant, je vais découvrir cette valorisation. En outre, la reconnaissance de cette valorisation va me permettre d'agir afin, par exemple, d'écouter plus cette musique. Mais rien de tout cela ne permet de réfuter que je la valorisais si l'on adopte la conception simple et donc strictement dispositionnelle de la valorisation. Un subjectivisme faisant appel à la conception simple de la valorisation fait donc mieux qu'une approche en termes de désirs.

Ce dernier point souligne au passage à quel point le caractère strictement dispositionnel de la conception simple de la valorisation la distingue de ce que la notion recouvre ordinairement. En effet, lorsque nous disons que nous valorisons quelque chose, alors cela implique que nous sachions ce qui est valorisé. Pourtant, dans le cas du massage et peut-être encore plus dans le cas du raga indien, nous ne savons pas que nous valorisons ces choses avant d'en faire l'expérience. Il s'ensuit que la valorisation entendue ordinairement subit comme le satisfactionnisme l'objection des plaisirs non anticipés. Mais, comme on vient de le remarquer au paragraphe précédent, le caractère strictement dispositionnel des valorisations selon la conception simple permet d'éviter cette objection et même d'y répondre.

Est-ce que cela rend douteuse ou mystérieuse la valorisation selon la conception simple ? Je ne le crois pas. Certes, il nous arrivera de ne pas savoir par avance ce qui sera susceptible de nous émouvoir et de nous apporter du plaisir, et donc ce qui sera susceptible de contribuer à notre bien-être. Mais cette difficulté épistémique n'enlève rien à la réalité ontologique de nos dispositions à apprécier telles ou telles choses. En outre, on peut parfois décrire, ne serait-ce qu'approximativement, le genre de choses qu'on est disposé à apprécier à un certain moment de sa vie. On notera en outre que l'appel au désir pose un problème équivalent. Il nous arrive de dire qu'on a envie de manger quelque chose de bon entendant par là quelque chose qu'on apprécierait. On peut dans une certaine mesure s'appuyer sur ce qu'on a aimé dans le passé, sur le genre de choses que nous apprécions, ce qu'on appelle parfois notre goût, pour éviter des déceptions trop importantes, mais nous restons incapables de savoir précisément quelles sont toutes les choses que nous pourrions aimer manger à ce moment. Il n'en

reste pas moins que nous avons à ce moment des dispositions qui déterminent complètement ce qu'on pourrait aimer. Et la thèse défendue ici est que ce sont ces dispositions, inconnues ou partiellement connues, qui déterminent ce qui peut contribuer à notre bien-être. Le fait que nous ne connaissions pas complètement ces dispositions n'enlève rien à la thèse subjectiviste que je propose ici.

Revenons maintenant à l'adéquation extensionnelle d'un subjectivisme de la valorisation conçue simplement. Un tel subjectivisme offre-t-il une meilleure adéquation extensionnelle que des subjectivismes faisant appel à des valorisations comprises de façon plus complexe ? Considérons le subjectivisme proposé par Tiberius. Il s'agit bien là d'une théorie subjectiviste du bien-être en termes de valorisation, comme la première partie du titre de son livre l'indique : *Well-being as value fulfillment : how we can help each other to live well* (2018). La conception de la valorisation (*to value*) qu'elle propose intègre les traits de la conception simple. Elle écrit en effet : « valoriser quelque chose dans son sens le plus abouti, c'est avoir un *pattern* relativement stable de dispositions ou tendances émotionnelles, motivationnelles et cognitives vis-à-vis de ce qui est valorisé » (2018, p. 10). Sa conception de la valorisation est toutefois plus complexe que la conception simple, et ce, en vertu de sa dimension cognitive. Elle précise : « valoriser quelque chose n'est pas seulement préférer cette chose ou l'apprécier, mais c'est aussi juger que c'est le genre de chose qui est bon d'une certaine manière » (p. 37).¹⁷ Plus loin, elle explique que cette condition cognitive revient finalement à introduire une condition d'assomption. Cependant, cette condition cognitive d'assomption soulève à nouveau l'objection posée par le cas de l'homosexuelle malgré elle.¹⁸ En effet, son orientation sexuelle contribuera *pro-tanto* à son bien-être alors que celle-ci ne sera pas valorisée au sens où Tiberius entend cette notion.

Jaworska propose une autre analyse de la pro-attitude de valorisation (*caring*) dans le cadre d'une théorie de l'agentivité. Celle-ci rejette la condition cognitive proposée par Tiberius pour les raisons que nous venons d'évoquer (2007, p. 537). Jaworska affirme toutefois que seules *certaines* dispositions émotionnelles manifestent une valorisation. Selon elle, des émotions comme la peur ou le dégoût ou des émotions qui sont liées à des désirs simplement passagers ne manifestent pas l'identité d'une personne (Jaworska, 2007, p. 555). Cependant, dans la mesure où notre perspective ici n'est pas de comprendre à quelles conditions une action peut être légitimement attribuée à un agent, mais de comprendre ce qui contribue au bien-être d'un individu, rien ne justifie d'adopter cette restriction. En effet, la douleur que nous éprouvons en nous cognant la tête sur un placard dans la cuisine affecte certainement notre bien-être, au moins au moment où nous souffrons même si elle ne définit nullement notre identité.

En résumé, les conditions supplémentaires sur les valorisations que nous avons considérées, qu'elles soient cognitives ou conatives, conduisent à nier à tort que

l'obtention de certaines choses contribue au bien-être. Le subjectivisme s'appuyant sur la valorisation conçue simplement est donc préférable du point de vue de l'adéquation extensionnelle aux subjectivismes sophistiqués, ou tout au moins ceux que nous avons considérés. Bien sûr, ceci n'est pas suffisant pour réfuter toutes les formes de subjectivismes qui pourraient faire appel à une valorisation comprise de façon plus complexe. Cependant, cela montre au moins les problèmes auxquels sont potentiellement soumises de telles versions du subjectivisme. Il en ressort que le subjectivisme faisant appel à une conception simple de la valorisation l'emporte au moins à ce stade de la discussion.

Notons au passage que cela laisse ouverte la possibilité pour le subjectivisme de la valorisation simple d'introduire une clause idéalisante. En effet, mon argument contre l'introduction d'une clause de cohérence ou de centralité dans le satisfactionnisme était que ces clauses ne permettent pas de régler l'objection des plaisirs non anticipés. Or, dans la mesure où le subjectivisme adoptant une conception simple de la valorisation évite cette objection, je n'ai rien dit jusqu'ici qui implique qu'il faudrait rejeter un subjectivisme de la valorisation simple avec une clause de cohérence ou de centralité. De façon similaire, dans la mesure où j'ai rejeté le satisfactionnisme avec une clause d'information idéale faible uniquement parce qu'il tombait sous l'objection des plaisirs non anticipés, rien n'interdit à première vue à un subjectivisme de la valorisation simple d'introduire une telle condition d'information idéale. Bref, la question de savoir si de telles clauses doivent ou non être introduites reste entière et ouverte. Il me faut toutefois la laisser de côté, car sa discussion m'entraînerait trop loin et m'écarterait de l'objet du présent article qui concerne uniquement la pro-attitude que le subjectivisme doit adopter.

3.3 Invariabilité

Il ne me reste donc plus qu'à évaluer dans quelle mesure un subjectivisme faisant appel à une pro-attitude de valorisation satisferait le desideratum d'invariabilité. En premier lieu, on peut noter à la suite de Lin qu'en faisant appel à une conception sophistiquée de la valorisation, le subjectivisme risque d'être incompatible avec le desideratum d'invariabilité. Cette critique s'applique pleinement à la conception de la valorisation adoptée par Tiberius. En effet, dans la mesure où la valorisation requiert qu'une personne possède des croyances sur ce qui est bon pour elle, il s'ensuit que les nourrissons ne valorisent rien ou qu'il faudrait leur appliquer une autre conception du bien-être.

L'introduction d'une condition strictement conative sur les valorisations pourra conduire à une objection semblable. Supposons que l'on utilise la conception de Jaworska de la valorisation pour soutenir que les objets valorisés simplement ne sont bons pour une personne que si, en outre, ces valorisations manifestent sa subjectivité particulière. Dans ce cas, parce que la valorisation négative des

objets qui suscitent ordinairement la peur ou le dégoût ne manifeste pas une subjectivité particulière, il s'ensuit que ces valorisations ne contribuent pas au bien-être d'un individu. Pourtant, il est assez plausible que des nourrissons ne possèdent pas d'autres formes de valorisation. Il s'ensuit qu'un subjectivisme qui adopterait cette conception de la valorisation ne satisfera pas le desideratum d'invariabilité.

Par contre, il est aisé de voir qu'un subjectivisme faisant appel à la conception simple de la valorisation satisfait le desideratum d'invariabilité,¹⁹ et ce, parce que la valorisation conçue simplement ne requiert aucune faculté qui ne serait pas possédée par un enfant à la naissance. Même si un nouveau-né ne possède pas la variété d'émotions d'un enfant plus âgé ou d'une personne adulte, il a indiscutablement des réactions émotionnelles, par exemple de peur, ou d'excitation, qui sont accompagnées de motivations, même si ces dernières sont très rudimentaires, comme le désir de téter le sein ou le biberon qu'il voit. Ainsi, pouvons-nous affirmer que l'enfant à la naissance valorise de nombreuses choses comme son alimentation, les soins qui lui sont prodigués, certains types d'environnements et d'interactions, etc. Un subjectivisme qui s'appuie sur la conception simple de valorisation satisfait donc le desideratum d'invariabilité. Un subjectivisme faisant appel à la conception simple de la valorisation est ainsi plus à même de satisfaire le desideratum d'invariabilité que des subjectivismes faisant appel à des conceptions plus sophistiquées de la valorisation. Mais bien sûr, il faut reconnaître que notre discussion ne permet pas d'exclure que des versions de la valorisation de sophistication intermédiaire puissent satisfaire le desideratum d'invariabilité.

J'en conclus qu'un subjectivisme qui ferait appel à la conception simple de la valorisation satisfait les trois desiderata que nous avons initialement défendus contrairement à au moins certaines formes de subjectivismes qui feraient appel à une notion de valorisation entendue en un sens plus riche. Comme la section 2 montrait que le subjectivisme satisfactionniste, qu'il soit simple, sophistiqué ou encore cognitif, échouait à satisfaire ces desiderata, il me semble que nous avons là un ensemble de raisons qui nous permettent d'affirmer que le subjectivisme devrait adopter la pro-attitude de valorisation comprise dans sa conception simple.

4. SUBJECTIVISME OU THÉORIE HYBRIDE ?

Pour finir, je voudrais évaluer dans quelle mesure l'appel à la pro-attitude de valorisation pourrait être une menace pour le subjectivisme et, à l'inverse, un argument en faveur d'une conception objective hybride du bien-être.

Une telle interrogation est motivée parce que de nombreux auteurs — tout au moins philosophes — jugent que les émotions sont par nature évaluables pour

leur adéquation, ou, autrement dit, ont des conditions de correction. Ces auteurs soutiennent en effet que les émotions — comme les croyances — sont, du fait de leur nature, susceptibles d’être correctes ou incorrectes.²⁰ Or, s’il en est ainsi, alors les valorisations elles-mêmes semblent évaluables pour leur correction de façon indirecte. En effet, si une émotion est correcte, alors on peut soutenir que la valorisation qui nous dispose à avoir cette émotion est elle-même correcte.

La question que nous devons par conséquent nous poser est celle-ci : ne faudrait-il pas admettre que le bien-être d’une personne ne se mesure pas tant à l’obtention de ce qu’elle valorise qu’à l’obtention de ce qui est valorisé *de façon appropriée ou adéquate* ? Si c’était le cas, alors cette position serait de toute évidence une position objectiviste hybride. Rappelons-le, une telle position objectiviste hybride combine deux thèses. D’un côté, elle affirme qu’un objet ne peut être bon pour une personne que si cet objet est intrinsèquement bon. Par exemple, une activité ludique ne peut être bonne pour une personne que si l’activité est intrinsèquement bonne. Mais, d’un autre côté, cette même activité ne peut être bonne pour cette personne que si cette activité ne lui est pas indifférente, bref si elle possède une certaine pro-attitude pertinente à l’égard de cette activité (Fletcher, 2013).²¹

En un mot, la menace semble donc la suivante : si nous défendons un subjectivisme qui fait appel à la notion de valorisation et si les valorisations ont des conditions d’adéquation au moins indirectes, le subjectiviste ne serait-il pas acculé à renoncer à sa position pour une position hybride ? Dans le cadre du présent article, il ne nous sera malheureusement possible ni de bloquer complètement cette objection ni de montrer à l’inverse qu’elle réfute le subjectivisme. Je souhaite toutefois offrir quelques considérations préliminaires montrant que le subjectivisme de la valorisation peut résister à la pente qui le ramènerait à un objectivisme hybride.²²

En premier lieu, le subjectivisme pourrait nier que les émotions ont des conditions d’adéquation (Kauppinen, 2014 ; Dokic et Lemaire, 2013, 2015 ; Schroeder, 2012). Après tout, la thèse selon laquelle le désir présente son objet comme bon et celle selon laquelle les désirs ont des conditions d’adéquation sont assez largement rejetées.²³ Pourquoi ne pourrait-on pas étendre cette thèse aux émotions ?

Une autre possibilité serait de soutenir que les émotions ont des conditions d’adéquation, mais que celles-ci sont relatives à nos valorisations.²⁴ Dans ce cas, l’émotion d’admiration qu’un individu éprouve face à un objet est adéquate relativement aux valorisations particulières de cet individu. Mais, dans la mesure où ses valorisations expliquent son admiration, la valorisation ne peut pas être inadéquate, puisqu’elle ne peut pas être inadéquate relativement à elle-même. Il s’ensuit que le théoricien qui soutiendrait que seules les valorisations adéquates

en ce sens contribuent au bien-être, dirait en réalité que seules les valorisations d'un individu contribuent à son bien-être. Or ceci n'est rien d'autre que la position subjectiviste. Bref, les conditions d'adéquation des émotions n'introduiraient pas une contrainte objective sur les objets valorisés qui contribueraient au bien-être.

Une troisième option serait de soutenir que les émotions ont des conditions d'adéquation objectives, tout en niant que seuls les objets valorisés de façon adéquate sont bons pour une personne. Après tout, rien ne semble interdire de soutenir d'un côté que l'œuvre de Bach est objectivement admirable, et que c'est là un fait objectif, tout en maintenant de l'autre qu'il ne s'ensuit pas que l'œuvre de Bach contribue positivement à mon bien-être.²⁵ Même si on définit la valeur objective comme la réponse que des experts auraient dans telles ou telles circonstances plus ou moins idéales, il se pourrait, pour moi qui ne suis pas un expert, que d'autres œuvres y contribuent même si les experts s'accordent à trouver médiocres ces autres oeuvres. Après tout, le concept de *bien* et celui de *bien pour* n'ont aucune raison d'être alignés, tout au moins faute d'argument supplémentaire.

En résumé, même si l'appel aux émotions peut éventuellement être le point de départ d'un argument en faveur d'une théorie hybride du bonheur, celui-ci ne nous conduit aucunement de façon nécessaire à une position hybride plutôt que strictement subjectiviste, et ceci même si les émotions ont des conditions d'adéquation. Bref, que le subjectivisme fasse appel à la valorisation comme pro-attitude pertinente ne le conduit pas malgré lui à une réfutation de sa propre théorie. Le débat qui oppose subjectivistes, objectivistes et théoriciens hybrides ne sera pas clos du fait d'adopter un subjectivisme de la valorisation et bien d'autres arguments devront être apportés pour le trancher.

5. CONCLUSION

J'ai initialement souligné que le désir a longtemps été considéré comme la seule pro-attitude permettant de défendre une approche subjectiviste du bien-être. Plus récemment, d'autres suggestions ont été faites. Dans le présent article, j'ai soutenu qu'au moins trois desiderata (celui d'adéquation extensionnelle, celui d'existence et celui d'invariabilité) contraignent le choix de la pro-attitude que le subjectivisme devrait adopter. L'hypothèse que je me suis efforcé de défendre est que seule une pro-attitude de valorisation conçue comme une disposition à éprouver des émotions ou du plaisir et à avoir des désirs relativement à un objet est capable de satisfaire ces desiderata. Le désir est un état trop transitoire pour expliquer les plaisirs et les émotions qui ne sont pas anticipés. Quant à la plupart des autres options, et en particulier les conceptions plus complexes de la valorisation, elles semblent conduire à une inadéquation extensionnelle qui ne peut être corrigée par l'introduction de conditions d'idéalisation. En outre, elles

sont, au moins pour celles qui s'écartent le plus notablement de la valorisation conçue simplement, incapables de satisfaire le desideratum d'invariabilité. Il apparaît donc que si une pro-attitude est constitutive des objets qui contribuent au bien personnel, alors cette pro-attitude est plausiblement la valorisation selon la conception simple que j'en ai proposée.

NOTES

- ¹ La question de savoir si les termes français *bonheur* et *bien-être* se recouvrent est disputée. Certains soutiennent qu'il faut les distinguer dans la mesure où une personne peut être heureuse alors que sa vie ne va pas bien. L'exemple standard est celui d'une personne qui est heureuse alors que ce bonheur repose sur des croyances largement fausses concernant par exemple l'estime ou l'affection que lui vouent ses proches (voir p. ex. Kagan, 1994). Dans la mesure où je ne souhaite pas prendre part ici à ce débat, je me contenterai d'utiliser le terme *bien-être*, laissant la question que je viens de présenter ouverte. Tout ce que je dirai sera donc compatible avec l'acceptation ou le rejet de cette distinction.
- ² Il faudrait dire « initiale au 20^e siècle », car en réalité on peut aisément rattacher ce type d'approche à Hobbes et Spinoza, entre autres.
- ³ Voir en particulier Kraut (2009) pour une défense de cette idée.
- ⁴ Il est plausible qu'il faille défendre une conception du bien-être qui s'applique aussi bien aux êtres humains qu'aux animaux doués de conscience, mais cette thèse ne nous intéressera pas ici. Je note toutefois qu'un certain nombre de positions subjectivistes, par exemple celles évoquant les désirs, sont susceptibles de satisfaire cette thèse plus ambitieuse.
- ⁵ Selon Dorsey (2017b), les capacités qui permettent à un individu à un certain âge de déterminer ce qui pour lui a de la valeur change complètement la perspective qu'on doit adopter sur son bien-être. Il me semble que cette thèse reste inacceptable car elle est métaphysiquement problématique. En effet, on ne peut justifier métaphysiquement comment une certaine relation entre un individu et une chose cesse d'exister du fait de se tenir en outre dans une relation nouvelle à une autre capacité. Une possibilité serait de répondre que nous devons comprendre le bien-être comme une propriété comparable à une institution qu'on peut décider de faire exister. Mais il est difficile d'assimiler le bien-être à une institution. En outre, même si c'était le cas, on ne voit pas pourquoi le bien-être pour ainsi dire « pré-institutionnel » disparaîtrait du fait de l'émergence d'un nouveau bien-être institué.
- ⁶ Je soutiens cependant dans Lemaire (soumis), que les attitudes que j'appelle ici de valorisation sont en mesure de résoudre ce problème.
- ⁷ Notez que si ces désirs peuvent être des désirs de second ordre, voire d'ordres supérieurs, il n'est pas requis — comme pour des formes de satisfactionnisme plus sophistiquées qu'on considérera plus bas — que les désirs soient par exemple eux-mêmes l'objet de désirs de deuxième ordre.
- ⁸ Dans son article contre le subjectivisme, Lin (2017, p. 362) reconnaît d'ailleurs qu'un subjectivisme faisant appel au désir satisfait l'exigence d'invariabilité. S'il critique le satisfactionnisme, c'est en réutilisant une version de l'argument de la machine à expérience de Nozick appliquée au nourrisson. Pour cela, il soutient (i) que le nourrisson désire uniquement des sensations, mais que (ii) sa vie serait moins bonne si ces sensations étaient produites par une machine à expérience. Je ne peux complètement ouvrir cette discussion dans le cadre du présent article. Notons toutefois que le subjectiviste pourra offrir ses réponses classiques consistant à nier soit la première soit la seconde hypothèse, voire les deux.
- ⁹ Je tiens à remercier un des lecteurs de la revue pour m'avoir poussé à discuter explicitement cette autre réponse.
- ¹⁰ Voir Bruckner (2016) dans le contexte d'une théorie du bien-être, mais aussi Lewis (1989) dans le cadre d'une théorie de la valeur ou encore Frankfurt (1971) et Dworkin (1988) dans le cadre d'une théorie de l'autonomie.
- ¹¹ Bien sûr, on pourrait abandonner la condition de résonance, mais l'abandonner serait précisément abandonner la considération centrale qui justifie les approches subjectivistes en général.
- ¹² L'approche proposée par Darwall (2002), selon laquelle ce qui est bon pour un individu est ce qu'une personne idéalement informée qui se soucierait de cet individu désirerait pour elle, aurait potentiellement les mêmes avantages, mais soulèverait aussi la présente objection.

- ¹³ Certes, on pourrait introduire une certaine dose d'idéalisation et en particulier d'idéalisation informationnelle. Cela pourrait conduire la personne en question à reconnaître qu'en réalité une sexualité homosexuelle serait bonne pour elle. Malheureusement, cette variante sera à nouveau confrontée aux mêmes problèmes liés à l'idéalisation que nous avons soulevés plus haut dans la section 2.3.
- ¹⁴ Comme le soutiennent par exemple les défenseurs des émotions comme programmes affectifs. Voir par exemple Panksepp (1982).
- ¹⁵ Cela ne préjuge nullement de la question de savoir si ces états mentaux sont réductibles ou non. On notera ici que la plupart des auteurs qui s'intéressent à la valorisation les conçoivent comme des attitudes complexes, possiblement composées d'éléments conatifs et cognitifs. Je souhaite toutefois rester neutre sur cette question à ce stade.
- ¹⁶ Je remercie un des évaluateurs du journal d'avoir soulevé cette objection.
- ¹⁷ Voir aussi Seidman (2009) pour l'introduction d'une condition cognitive proche dans le cadre d'une théorie de l'agentivité et de l'identité subjective.
- ¹⁸ La critique que je formule ici repose sur une interprétation littérale des thèses qu'elle soutient dans son livre. Il me semble toutefois qu'elle décrit elle-même de façon inadéquate la position qui semble la conclusion de certains de ses arguments. En effet, lorsqu'elle insiste sur la pertinence de l'endossement ou de la cohérence des valorisations, ses arguments consistent essentiellement à dire que les valorisations endossées et cohérentes seront plus faciles à satisfaire conjointement, ce qui permettra ainsi d'atteindre un niveau de bien-être supérieur. Mais si c'est là son argumentation, alors ces conditions ne devraient pas entrer dans la définition de ce qui contribue au bien-être. Il faudrait maintenir que seule la satisfaction des valorisations compte tout en soulignant que certains ensembles de valorisations, du fait d'être plus cohérents et d'être endossés, sont plus susceptibles d'être conjointement satisfaits et donc de produire un niveau de bien-être supérieur.
- ¹⁹ Lin soutient qu'un subjectivisme de la valorisation n'est pas en mesure de satisfaire le desideratum d'invariabilité (2017, p. 361). Il faut toutefois noter qu'il range sous cette catégorie des conceptions qui en réalité font appel à des désirs et que j'ai qualifié de sophistiquées. Il ne considère pas l'approche de Tiberius, ou la possibilité d'une valorisation simple. S'il rejetait la conception de Tiberius pour les raisons présentées dans le texte, il me semble qu'il devrait concéder, comme pour le satisfactionnisme, que ces formes de subjectivismes sont compatibles avec la thèse d'invariabilité. En effet, et comme on l'a vu plus haut, les objections de Lin contre ces formes de subjectivisme portent alors sur des points totalement distincts, et que nous ne pouvons discuter ici.
- ²⁰ Ces auteurs soutiennent par exemple qu'une émotion comme la peur est correcte ou appropriée vis-à-vis des objets dangereux ou effrayants. Ils offrent ensuite différentes analyses de la notion de correction.
- ²¹ Si c'est là l'interprétation la plus courante de ce qu'est une théorie hybride du bien-être, d'autres façons d'hybrider subjectivisme et objectivisme sont concevables. Voir par exemple Wall et Sobel (2021).
- ²² Je remercie Christine Tappolet d'avoir soulevé cette importante question.
- ²³ Même si elle est défendue sous différentes versions par Tenenbaum (2007), Gregory (2013), Orsi (2015), Lauria (2017).
- ²⁴ Cette position me paraît largement compatible avec l'approche de Helm (2001). Lépine (2016) et Müller (2019) défendent des versions proches de cette idée dans laquelle la pro-attitude est plutôt un désir qu'une valorisation.
- ²⁵ Sans adopter cette position, Levinson (2010) envisage la possibilité d'un tel écart. Sa position est néanmoins intéressante, car il ne soutient pas la thèse du théoricien hybride mais une position intermédiaire selon laquelle, la plupart du temps et dans de bonnes conditions, écouter une musique qui est objectivement bonne contribue plutôt à notre bien-être.

RÉFÉRENCES

Arneson, Richard J., « Human Flourishing Versus Desire Satisfaction », *Social Philosophy and Policy*, vol. 16, no. 1, 1999, p. 113-142.

Brandt, Richard B., *A Theory of the Good and the Right*, 2^e éd., Oxford, Clarendon Press, 1998.

Bruckner, Donald W., « Quirky Desires and Well-Being », *Journal of Ethics & Social Philosophy*, vol. 10, no. 2, 2016, p. 1-34.

Darwall, Stephen, *Welfare and rational care*, Princeton, Princeton University Press, 2002.

Dokic, Jérôme, et Stéphane Lemaire, « Are Emotions Perceptions of Value? » *Canadian Journal of Philosophy*, vol. 43, no. 2, 2013, p. 227-247.

———, « Are Emotions Evaluative modes? », *dialectica*, vol. 69, no. 3, 2015, p. 271-292.

Dorsey, Dale, « Subjectivism without Desire », *Philosophical Review*, vol. 121, no. 3, 2012, p. 407-442.

———, « Why Should Welfare 'Fit'? », *The Philosophical Quarterly*, vol. 67, no. 269, 2017, p. 685-624.

Dworkin, Gerald, *The Theory and Practice of Autonomy*, New York, Cambridge University Press, 1988.

Ellsworth, Phoebe C., et Klaus R. Scherer, *Appraisal Processes in Emotion*, Oxford University Press, 2001.

Fletcher, Guy, « A Fresh Start for the Objective-List Theory of Well-Being », *Utilitas*, vol. 25, no. 2, 2013, p. 206-220.

Frijda, Nico H., *The Laws of Emotion*, New York, Psychology Press, 2007.

Frankfurt, Harry, « Freedom of the Will and the Concept of a Person », *The Journal of Philosophy*, vol. 68, no. 1, 1971, p. 5-20.

Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, MA, Harvard University Press, 1990.

Gregory, Alex, « The Guise of Reasons », *American Philosophical Quarterly*, vol. 50, no. 1, 2013, p. 63-72.

Heathwood, Chris, « The Problem of Defective Desires », *Australasian Journal of Philosophy*, vol. 83, no. 4, 2005, p. 487-504.

———, « Desire Satisfactionism and Hedonism », *Philosophical Studies*, vol. 128, no. 3 2006, p. 539-563.

———, « Which desires are relevant to well-being? » *Noûs* vol. 53, no. 3, 2019, p. 664-688.

Helm, Bennett W., *Emotional Reason: Deliberation, Motivation, and the Nature of Value*, Cambridge, Cambridge University Press, 2001.

Jaworska, Agnieszka, « Caring and Internality », *Philosophy and Phenomenological Research*, vol. 74, no. 3, 2007, p. 529-568.

Kagan, Shelly, « Me and My Life », *Proceedings of the Aristotelian Society*, vol. 94, 1994, p. 309-324.

Kauppinen, Antti, « Fittingness and Idealization », *Ethics*, vol. 124, no. 3, 2014, p. 572-588.

Kraut, Richard, *What Is Good and Why: The Ethics of Well-Being*, Cambridge, MA, Harvard University Press, 2009.

Lauria, Federico, « The “Guisse of the Ought to Be”: A Deontic View of the Intentionality of Desire », in Federico Lauria et Julien A. Deonna (dir.), *The Nature of Desire*, New York, Oxford University Press, 2017, p. 139-164.

Lemaire, Stéphane, « Subjectivism and the Explanatory Challenge », manuscrit.

Lepine, Samuel, « Les découvertes émotionnelles et les motivations affectives », *Revue philosophique de la France et de l'étranger*, vol. 141, no. 2, 2016, p. 179-196.

Levinson, Jerrold, « Artistic Worth and Personal Taste », *The Journal of Aesthetics and Art Criticism*, vol. 68, no. 3, 2010, p. 225-233.

Lewis, David, « Dispositional Theories of Value », *Proceedings of the Aristotelian Society, Supplementary Volumes*, vol. 63, 1989, p. 89-174.

Lin, Eden, « Against Welfare Subjectivism », *Noûs* vol. 51, no. 2, 2017, p. 354-377.

———, « Welfare Invariabilism », *Ethics*, vol. 128, no. 2, 2018, p. 320-345.

Müller, Jean Moritz, *The World-Directedness of Emotional Feeling: On Affect and Intentionality*, Cham, Palgrave Pivot, 2019.

Orsi, Francesco, « The Guise of the Good », *Philosophy Compass* vol. 10, no. 10, 2015, p. 714-724.

Ortony, Andrew, Gerald L. Clore et Allan Collins, *The Cognitive Structure of Emotions*, Cambridge, Cambridge University Press, 1990.

Panksepp, Jaak, « Toward a General Psychobiological Theory of Emotions », *Behavioral and Brain Sciences*, vol. 5, no. 3 (1982), p. 407-422.

Railton, Peter, « Moral Realism », *The Philosophical Review*, vol. 95, no. 2, 1986a, p. 163-207.

———, « Facts and Values », *Philosophical Topics*, vol. 14, n° 2, 1986b, p. 5-31.

Rosati, Connie S., « Persons, Perspectives, and Full Information Accounts of the Good », *Ethics*, vol. 105, no. 2, 1995, p. 296-325.

———, « Personal Good », in Terry Horgan et Mark C. Timmons, *Metaethics After Moore*, Oxford, Oxford University Press, 2006, p. 107-132.

Scarantino, Andrea, « The Motivational Theory of Emotions », in Justin D'Arms et Daniel Jacobson (dir.), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*, Oxford, Oxford University Press, 2014, p. 156-185.

Schroeder, Mark, « The Ubiquity of State-Given Reasons », *Ethics*, vol. 122, no. 3, 2012, p. 457-488.

Seidman, Jeffrey, « Valuing and Caring », *Theoria*, vol. 75 no. 4, 2009, p. 272-303.

Sidgwick, Henri, *The Methods of Ethics*, 7^e éd., Cambridge, MA, Harvard University Press, 1907.

Smith, Michael, « Internal Reasons », *Philosophy and Phenomenological Research*, vol. 55, no. 1, 1995, p. 109-131.

Sobel, David, « Full Information Accounts of Well-Being », *Ethics*, vol. 104, no. 4, 1994, p. 784-810.

———, « Subjectivism and Idealization », *Ethics*, vol. 119, no. 2, 2009, p. 336-352.

Tappolet, Christine, *Emotions, Values, and Agency*, New York, Oxford University Press, 2016.

Tiberius, Valerie, *Well-being as Value Fulfillment: How We Can Help Each Other to Live Well*, New York, Oxford University Press, 2018.

Tenenbaum, Sergio, *Appearances of the Good: An Essay on the Nature of Practical Reason*, Cambridge, Cambridge University Press, 2007.

Wall, Steven et David Sobel, « A Robust Hybrid Theory of Well-Being », *Philosophical Studies*, vol. 178, no. 9, 2021, p. 2829-2851.

BIEN-ÊTRE, AFFECTIVITÉ ET SOCIÉTÉ : ENJEUX MORAUX ET ENJEUX STRUCTURAUX

JULIEN CLAPARÈDE-PETITPIERRE

UNIVERSITÉ CLERMONT-AUVERGNE - PHIER

RÉSUMÉ :

Cet article aborde la question du bien-être à partir du problème sociologique et psychologique de la socialisation. Deux types de théories distinctes de la socialisation sont ici cernés qui posent de façon différente la question du rapport entre émotions et bien-être. En premier lieu, les théories de la répression (Freud) et des sentiments moraux (Elster) font du processus de socialisation une expérience d'émotions négatives puissantes suscitées par l'intériorisation du jugement moral d'autrui dans la psyché individuelle. Qu'elle soit naturalisante (Freud) ou culturaliste (Benedict, Elster), cette approche repose sur l'idée d'une contradiction entre les aspirations pulsionnelles de l'individu et les réquisits moraux du groupe. Cette contradiction, qui induit une violence affective (angoisse du rejet, honte) et un renoncement systématique aux satisfactions du désir, montre la socialisation comme un processus conflictuel générant inévitablement du mal-être. En second lieu, les théories de l'apprentissage font de la socialisation un processus générateur de bien-être dans la mesure où la découverte cognitive des règles qui structurent le champ social permet à l'individu de devenir un acteur social compétent et à l'aise dans les interactions quotidiennes. Ces théories distinguent l'affectivité de haute intensité, générée par l'angoisse ressentie par l'acteur social incompetent et désorienté, de l'affectivité de basse intensité reposant sur le sentiment de confiance affectant l'acteur social compétent (Quéré). Cette seconde approche illustre l'intérêt que porte l'individu à la maîtrise des règles sociales. Quoique distinctes, ces deux positions ne s'opposent pas, voire semblent pouvoir être articulées à profit.

ABSTRACT:

This paper addresses the issue of well-being in the overall context of socialization. It pinpoints two distinct theories of socialization, each of which induces a specific articulation between emotions and well-being. According to Freud's theory of repression and to Elster's theory of moral sentiments, the socialization process is regarded as an experience of powerful negative emotions triggered by the internalization of moral judgment. Whether naturalistic (Freud) or culturalist (Benedict, Elster), this approach is based on the idea of a contradiction between the individual's impulsive aspirations and the moral demands of the group. This contradiction induces an affective violence (anguish, shame) and a systematic renunciation to the realization of desires. It depicts socialization as a conflictual process generating an inevitable malaise. In contrast, learning theories see socialization as a possible source of well-being. Indeed, the cognitive elucidation of social rules allows the individual to become a competent social actor and comfortable in daily

interactions. These theories distinguish between a high-intensity affectivity generated by the anxiety felt by the incompetent and disoriented social actor, and a low-intensity affectivity based on the feeling of confidence affecting the competent social actor (Quéré). This second approach illustrates the individual's interest in mastering social rules. Although distinct, these two positions are not opposed to each other, and even seem to be able to be articulated to advantage.

INTRODUCTION

Il est important, quand on aborde la question des émotions et du bien-être de l'être humain, de le considérer en tant qu'être social. En effet, la société peut apparaître soit comme pourvoyeuse de frustrations, soit, inversement, comme l'espace au sein duquel l'individu peut s'épanouir émotionnellement. Afin de clarifier cette question, on s'intéressera ici à l'ambiguïté de la notion de « norme sociale ». En effet, on note que certains auteurs (Freud, 1981 [1920] ; Elster, 1998, 2008, par exemple) considèrent la normativité sociale principalement comme un phénomène moral impliquant la valorisation du bien et la condamnation du mal. De l'autre, on observe que, chez d'autres auteurs, la notion de norme sociale renvoie fréquemment à des rites d'interaction dénués de valeur immédiatement morales (Garfinkel, 1967 ou Goffman, 1967, par exemple). Si ces deux approches nous paraissent légitimes et sont susceptibles d'être articulées, il nous semble également nécessaire de les distinguer clairement. Ce faisant, nous verrons que la question du bien-être et de son rapport aux émotions se pose de façon très différente et implique des solutions tout aussi hétérogènes selon que l'on considère l'aspect moral ou l'aspect structural de la normativité sociale.

Ainsi, en revenant d'abord sur les travaux de Freud et de l'anthropologie culturelle, nous pourrions voir de quelle façon la normativité morale d'une société influe sur le bien-être des individus et nous décrirons le rôle des sanctions émotionnelles dans le processus d'établissement et d'intériorisation de la morale. Nous montrerons notamment que le bien-être individuel peut être évalué en fonction du degré de convergence ou de divergence existant entre les tendances psychiques des individus et les réquisits moraux de la culture locale. Nous reviendrons également sur le rôle des affects comme l'angoisse et la honte dans le processus d'intériorisation des normes morales et du jugement d'autrui.

Nous verrons par la suite que la normativité structurale, quoique moins coercitive que la normativité morale, est déterminante pour comprendre les conditions du bien-être de l'individu en société. Ce second questionnement impliquera de rompre avec l'approche morale pour privilégier un mode de questionnement propre à la psychologie de l'apprentissage. Il sera alors montré que la maîtrise cognitive des normes, règles et signes qui composent l'environnement social de l'individu génère un sentiment de familiarité qui est décisif pour son bien-être. Réciproquement, nous verrons que les expériences de désorientation tenant au non-respect des rites d'interactions normalisés dans les interactions quotidiennes conduisent à une situation d'anxiété délétère pour l'individu. Cette anxiété se présente, nous l'observerons, comme une tonalité affective de haute intensité induisant d'une part des blocages cognitifs et, d'autre part, une fébrilité produisant des difficultés à gérer ses émotions, notamment les émotions négatives comme la colère.

Dès lors, cette contribution s'inscrit dans le cadre du dialogue contemporain entre sciences sociales et philosophie en ce qui concerne la question du lien entre émotions et normativité sociale (Quéré, 2021). Une telle entreprise implique

notamment de chercher à articuler une approche structurale, au sens large, issue de la sociologie et de l'anthropologie sociale (Radcliffe-Brown, 1952 ; Garfinkel, 1967 ; Goffman, 1967) avec la psychologie des émotions (Saarni, 2022), la tradition psychanalytique et l'approche morale issue de la philosophie des sentiments moraux (Elster, 1998, 2008). Ce travail permet de mettre en lumière les spécificités d'une approche structurale du problème de la normativité et du bien-être en montrant que l'adoption d'une conduite normée répond à un intérêt cognitif dont la valorisation immédiate est le sentiment de confiance ressenti par tout individu confronté à un défi cognitif qu'il se sait capable de relever. Ce premier résultat conduit également à saisir que la satisfaction cognitive procure un confort émotionnel qui favorise en retour à fois la gestion des émotions et l'acuité cognitive. Ainsi, l'approche structurale et cognitive de la question de la normativité apparaît comme un complément nécessaire de l'approche morale puisque le sujet ne se conforme pas aux normes dans le seul intérêt de se sentir valorisé en tant qu'ego, mais aussi parce que toute situation sociale se présente comme une forme de jeu structural, c'est-à-dire de défi de type cognitif à relever afin de faire preuve de sa compétence d'acteur social.

I. LA NORMATIVITÉ SOCIALE INTERDIT-ELLE LE BIEN-ÊTRE INDIVIDUEL ? FRUSTRATION, ANGOISSE ET RÉPRESSION

A/ Freud : la socialisation comme endiguement des exigences pulsionnelles

Le premier modèle relevant d'une approche morale du problème de la normativité sociale que nous voulons rappeler est celui de Freud dans la mesure où sa théorie pulsionnelle de l'instinct a marqué historiquement la réflexion en sciences sociales, notamment en anthropologie¹. L'approche du médecin viennois est fondée sur l'idée que la psychologie humaine est animée de pulsions. Les pulsions (*Trieben*) sont, pour Freud (2016 [1929]), des désirs intenses et impérieux dont la satisfaction procure une joie qui constitue le degré maximal du bonheur. Le prototype de cette représentation est l'activité sexuelle². Selon cette conception du bonheur, le degré maximal de bien-être ne peut être obtenu que par la satisfaction immédiate, non empêchée, non reportée, non détournée et non contrainte des pulsions à mesure qu'elles sont ressenties. Ce degré maximal de bonheur se laisse donc décrire à partir d'un égocentrisme imposant ses exigences pulsionnelles. En effet, l'économie libre et spontanée de la vie pulsionnelle se caractérise, comme le désir sexuel, par la succession de phases aiguës de désir et de moments de satisfaction impliquant l'extinction temporaire du désir³.

La présentation du psychisme pulsionnel par Freud fait du bonheur maximal humain une succession de phases de tensions progressives et de décharges brutales réglées sur la périodicité du désir individuel. Cette vie guidée exclusivement par le principe de plaisir est incompatible avec la socialité et, selon Freud, le processus de civilisation réside dans la domestication des pulsions humaines. Dans ce nouveau cadre fait de normes et de tabous, l'être humain expérimente une exis-

tence diminuée reposant sur l'inhibition des pulsions et l'atténuation de l'amplitude émotionnelle, celle de la joie de la satisfaction comme celle de la frustration. Si les souffrances sont étouffées, les satisfactions sont de faible intensité⁴. Loin de considérer la tempérance comme pourvoyeuse de bien-être selon le modèle antique de la *sophrosunê* platonicienne ou de l'ataraxie épicurienne, Freud considère la stabilisation émotionnelle de la vie psychique humaine comme la source d'une insatisfaction permanente qui fait de l'être humain socialisé un névrosé chronique⁵. Le renoncement à la satisfaction directe des pulsions humaines nécessaire à l'harmonie sociale conduirait ainsi les êtres humains à vivre dans une frustration liée à une modération mal ajustée aux variations et aux urgences de la vie pulsionnelle. De là s'ensuivrait ce malaise dans la civilisation évoquée par Freud, c'est-à-dire ce renoncement pulsionnel auquel est contraint l'homme civilisé qui l'empêcherait de goûter au véritable bonheur.

Pensant le processus de civilisation comme un processus d'érosion des excès émotionnels, Freud décrit en creux une nature humaine faite d'exigences pulsionnelles impérieuses et amatrice d'émotions puissantes, contrastées et excessives. La police des mœurs viendrait écrêter l'amplitude émotionnelle de cette psychologie initiale en la domestiquant, c'est-à-dire en restreignant la satisfaction des désirs.

B/ Mécanisme émotionnel d'intériorisation des interdits sociaux : le sentiment d'angoisse

Si l'on suit la position freudienne, le bien-être de l'individu socialisé n'est pas seulement amoindri par l'insatisfaction liée à la répression, au moins partielle, des appétits pulsionnels. Il faut également prendre en compte le mécanisme par lequel s'impose l'interdit social. Ce mécanisme chargé d'endiguer les mouvements pulsionnels repose sur l'intériorisation des normes de la moralité rendue effective par l'érection de ce que Freud appelle le *Surmoi*. Or, l'efficacité de ce Surmoi repose sur un type d'émotion ressenti par l'individu : le sentiment d'angoisse. Selon Freud, l'angoisse est un sentiment primitif ressenti par tous les nourrissons humains qui sourd de la situation de fragilité radicale du nouveau-né. Profondément angoissé par la précarité de sa survie, le nourrisson est constamment hanté par la peur bouleversante de l'abandon. Cette tonalité émotionnelle première, puissante et omniprésente dans les premiers temps de la vie humaine, a pour effet de nourrir un désir aigu d'être aimé et protégé par une instance parentale. Par voie de conséquence, l'enfant apprend à renoncer à toute satisfaction pulsionnelle conduisant à une désapprobation majeure de la part de l'instance familiale protectrice. Cette instance familiale se faisant le relai de la normativité sociale morale au sein du foyer conduit l'enfant à intérioriser les normes sociales morales, c'est-à-dire à renoncer de lui-même à satisfaire certaines injonctions pulsionnelles en raison de l'angoisse du désamour et de l'abandon qui en résulterait.

Angoissé à l'idée de perdre l'amour et l'attention d'autrui, le sujet en vient à censurer tous les désirs susceptibles d'être désapprouvés par autrui⁶. Selon

Freud, le Surmoi s'érige depuis la capacité à anticiper la condamnation morale d'autrui et à ressentir l'angoisse de l'abandon qui s'ensuivrait. Se nourrissant du sentiment de vulnérabilité ressenti par l'enfant à la naissance, le Surmoi fait donc planer une angoisse latente et permanente sur la vie humaine. L'individu socialisé n'est donc pas seulement un individu dont la vie émotionnelle est écrêtée, c'est également un individu hanté par le sentiment aigu et violent d'angoisse. Le renoncement à la satisfaction pulsionnelle et à l'intensité d'une vie émotionnelle épanouie permet de calmer cette menace sourde qui habite constamment la vie psychologique de l'être humain. Ainsi, la théorie de Freud rend compte de l'intériorisation d'une culpabilité morale qui ne se manifeste pas seulement face à l'autorité ponctuelle d'autrui. La constitution du Surmoi place en permanence le sujet sous le regard accusateur et sévère d'un autrui devenu une instance intériorisée⁷. Si autrui est une figure dont on peut contourner la sévérité par la dissimulation, le Surmoi est en revanche une altérité intériorisée au regard de laquelle rien n'échappe. Ainsi, dès lors que l'individu cherche la satisfaction pulsionnelle en transgressant les restrictions sociales, l'omniscient Surmoi plonge le sujet dans les affres de l'angoisse. Le sujet psychologique se trouve contraint à un affadissement émotionnel de l'existence pour contenir la violence du sentiment d'angoisse toujours prompt à se réveiller⁸. Engagée dans une tragédie du désir fondée sur l'opposition entre les exigences de la vie pulsionnelle et celle de la socialité, la psychologie humaine se présente donc, selon Freud, comme un douloureux équilibre entre la frustration et l'angoisse.

II. APPROCHE CULTURALISTE DE LA QUESTION DE LA NORMATIVITÉ SOCIALE

A/ Réfutation de l'idée selon laquelle socialiser serait endiguer la pulsionnalité humaine

L'un des premiers textes de Gregory Bateson est consacré à la question des rapports entre la psychologie et l'ordre social. Il s'agit d'un article intitulé « Psychology and the War » publié dans le *Times* du 13 décembre 1934 dans lequel l'anthropologue britannique, nourri de sa lecture de Ruth Benedict (1960 [1934]), affirme l'importance d'une approche culturelle des mentalités. Répondant à un certain Pr William Brown, Bateson prend part à un débat portant sur la question de l'agressivité naturelle de l'être humain. Alors que l'Europe se trouve déjà, cinq ans avant la Seconde Guerre mondiale, dans une phase de tensions diplomatiques majeures, certains experts prétendent que la guerre est inévitable en raison des instincts agressifs propres à la nature humaine. Invoqué comme cause des guerres, cet instinct martial – comme la pulsion sexuelle décrite par Freud – serait un danger constant pour la paix des sociétés humaines⁹. Adoptant une position typiquement freudienne, Brown rappelle l'importance du mécanisme de sublimation – détournement de la pulsion vers des activités plus nobles comme l'art en garantissant ainsi son innocuité – quant à la gestion civilisée d'une pulsionnalité humaine antisociale.

La référence à une humanité primitive et sauvage dont les traits psychologiques pulsionnels excessifs seraient mal réprimés sous une socialité normée est un poncif auquel Bateson s'attaque en qualité d'anthropologue. Il explique que, loin de permettre la mise en relief des traits particuliers de l'humanité, l'anthropologie – en s'appuyant sur les descriptions ethnographiques d'un grand nombre de groupes humains – tend bien plutôt à démontrer la vacuité d'une telle prétention. La position de Bateson consiste dans le refus de considérer les phénomènes de normativité sociale sous l'angle strict de la répression. De même que la sexualité peut être culturellement traitée de façon variable par différentes sociétés, l'agressivité est également susceptible d'être inscrite de façon variable dans l'espace social en fonction du régime normatif de chaque culture¹⁰.

Loin de considérer la guerre comme irruption d'une brutalité pulsionnelle innée de l'homme sauvage dans l'espace pacifié de la civilisation, Bateson aborde la question avec la précaution du culturaliste. Rien n'autorise à penser que la guerre est l'effet d'une pulsion primaire d'agressivité qui échapperait aux stratégies sociales d'endiguement. Ce qui conduit l'anthropologue à douter de cette explication est le fait que les comportements agressifs ne sont pas rejetés par toutes les cultures : certains peuples laissent peu de place à la violence quand d'autres la valorisent largement. Bateson entend donc montrer que la pratique de la guerre n'est pas une brèche dans l'ordre social européen, mais une conséquence de celui-ci. Aussi, les modalités agonistiques européennes ne doivent pas être expliquées par l'irruption d'une pulsionnalité primitive venant rompre les digues de la normativité sociale. Ces tendances martiales sont au contraire le résultat et le symptôme d'un état culturel des sociétés occidentales durant cette période : ce sont précisément les cultures de ces sociétés qui suscitent et entretiennent une telle agressivité. L'erreur de la position freudienne, selon l'approche culturaliste, consiste dans la croyance selon laquelle violence et sexualité sont des comportements primaires qui, premièrement, n'appartiendraient pas à l'ordre social et, deuxièmement, lui seraient contraires. Bateson rappelle que dans certaines sociétés la rixe est inscrite au régime normal des interactions et constitue un facteur d'intégration sociale.

Rien n'oblige une société à réprimer la violence ou la sexualité. Ainsi, des tendances similaires – violence exacerbée ou sexualité intense – peuvent être valorisées ou dévalorisées selon les coordonnées culturelles. Dans *Naven* en 1936, Bateson montre que, dans la tribu de guerriers coupeurs de tête des Iatmuls de Nouvelle-Guinée, les comportements de violence et les affects de colère, voire de rage, sont très valorisés. Au contraire, les individus dont l'affectivité est plus mesurée sont méprisés pour leur lâcheté et leur pusillanimité. L'agressivité, avec son cortège d'émotions et de comportements spécifiques, peut être institutionnalisée de façon différente – notamment rejetée ou valorisée – selon les cultures. Bateson note ainsi que la similarité constatée entre les pratiques guerrières des Européens et celle d'autres peuples non européens comme les Lae Womba ou les Zulus ne doit pas être lue comme l'expression d'une pulsionnalité primitive, mais comme le résultat d'une proximité des options culturelles adoptées par ces sociétés quant aux modalités sociales du régime de la violence.

L'ordre social n'est donc ni conforme ni réfractaire aux caractéristiques pulsionnelles de la nature humaine et le rapport entre affectivité individuelle et normativité sociale ne doit pas être conçu de façon simpliste selon le doublet pulsion / répression.

B/ Configuration culturelle et psychologie individuelle : adaptation et standardisation des émotions

La même année, en 1934, Ruth Benedict publie *Patterns of Culture* dont le chapitre IV est consacré à l'analyse des rapports entre l'individu et la configuration culturelle. Tout comme Bateson, elle affirme que la société et l'individu ne sont pas antagonistes¹¹. L'opinion de « l'homme de la rue » selon laquelle la société est une instance de répression est un lieu commun duquel il faut se départir pour Benedict¹². La culture d'une société ne doit pas être vue comme une instance d'inhibition et de répression d'une pulsionnalité pourvoyeuse d'une vie émotionnelle intense : au contraire, nombre de satisfactions émotionnelles sont intimement liées à la configuration culturelle qui permet d'enrichir l'affectivité des membres d'une société. Reconsidérant cette question du traitement des émotions par l'anthropologie sociale et culturelle, Louis Quéré revient sur l'idée d'une « discipline sociale des émotions » (Quéré, 2021, p. 134) dont il retrace l'histoire et les linéaments par l'étude de Durkheim, Halbwachs, Mauss, Benedict et Karsenti. En rappelant qu'il existe des « répertoires émotionnels différents selon les cultures » (Quéré, 2021, p. 129), Quéré s'inscrit ici dans une tradition théorique qui remonte à la sociologie de Durkheim et Mauss ainsi qu'à l'anthropologie culturelle de Benedict.

Selon Benedict, les individus socialisés sont psychologiquement standardisés de deux façons par les configurations culturelles au sein desquelles ils se trouvent : ils sont d'une part *influencés* et d'autre part *sélectionnés*.

Le premier point renvoie à une discipline sociale des émotions ainsi que des comportements. Dans une société qui valorise la violence et une certaine forme de nervosité ambiante – comme chez les Iatmuls décrits par Bateson – les émotions de colère, de rage et d'aigreur ne seront ni réprimées ni dissimulées : au contraire, elles seront encouragées, exagérées, cultivées. À l'inverse certaines cultures – Bateson évoque le flegme de la haute société britannique ou l'attitude nonchalante des Balinais – imposeront une discipline des affects tout à fait opposée en faisant de la nervosité et de l'expression de la colère le signe de la perte de contrôle de soi et donc d'une certaine infériorité morale. Il ne faut pourtant pas se méprendre sur le sens qu'une approche culturaliste donne à ces deux options culturelles : il n'y en a pas une qui serait laxiste tandis que l'autre serait répressive. Ces deux modèles de gestion des émotions ne sont en soi ni répressifs ni favorables au déploiement d'une vie émotionnelle épanouie. En effet, selon les culturalistes il n'y a pas d'appétit spontané et universel de l'être humain pour les attitudes émotionnelles excessives. De surcroît, la culture et la normativité sociale ne vont pas toujours de pair avec un affadissement de l'amplitude émotionnelle ; cela peut être l'inverse et les individus peuvent être contraints

d'adopter une sensibilité extrême et d'afficher des passions violentes comme les Iatmuls décrits par Bateson. Or, puisqu'il n'y a pas d'appétit universel de l'être humain pour les attitudes émotionnelles extrêmes, cela signifie qu'une culture valorisant l'intensité émotionnelle peut être aussi coercitive qu'une culture favorisant la modération. Tout dépend des tendances spontanées de l'individu : un individu doux et introverti sera tout aussi malheureux chez les Iatmuls qu'un individu sanguin et extraverti dans un club britannique du début du XX^e siècle.

Pour Benedict, il faut donc considérer que tous les individus ne sont pas adaptés de façon égale à la configuration culturelle de la société au sein de laquelle ils naissent. En effet, si le caractère des individus est pour une large part *influencé* par la culture locale au gré du processus de socialisation des enfants, les individus sont également *sélectionnés* en fonction des convergences ou des divergences de leurs tendances psychologiques initiales avec la configuration culturelle de leur société. Ainsi, si la culture iatmule est peu coercitive pour les individus violents et extravertis, elle l'est beaucoup plus pour les individus doux et introvertis. On observe alors que les individus dont l'affectivité est la plus spontanément concordante avec ces réquisits normatifs sont appelés à devenir des chefs ou des personnes puissantes tandis que les autres sont rabaissés à des rôles subsidiaires et méprisés. Ainsi sont *sélectionnés* les individus en fonction de leurs capacités à épouser les normes affectives et comportementales locales.

Le fait qu'il existe toujours des individus inadaptés dans toutes les configurations culturelles recensées¹³ montre qu'il n'y en a pas une qui soit plus conforme aux réquisits pulsionnels de l'être humain : ces réquisits changent en fonction des individus et l'exacerbation de la violence ou de la sexualité ne sont pas les conditions universelles du bonheur individuel. Si l'on adopte la position de Benedict, il apparaît que Freud, loin de décrire la nature innée de tout être humain désireux de maximiser l'intensité des satisfactions pulsionnelles, montre que certains individus sont mal à l'aise avec la configuration culturelle qu'ils rencontrent dans l'Autriche de leur temps.

Dans *Patterns of Culture*, Benedict oppose ainsi les cultures dionysiaques – qui favorisent les comportements excessifs, une sexualité extravertie ainsi qu'un certain appétit pour l'intensité émotionnelle et la perte de contrôle de soi – aux cultures apolliniennes – favorisant la modération et la retenue des comportements affectifs ainsi que le contrôle de soi. La lecture de Benedict montre que si les cultures dionysiaques existent, elles n'en sont pas moins des cultures particulières. Cela signifie que l'exacerbation des conduites dites pulsionnelles relève autant d'une discipline culturelle des émotions que la valorisation de la modération. Autrement dit, exacerber l'impulsivité et les appétits sexuels n'est pas moins contraignant *en soi* que les restreindre : tout dépend des tendances spontanées de l'individu soumis à une telle discipline des émotions et des comportements.

C/ Mécanisme émotionnel d'intériorisation des interdits moraux d'une culture : le sentiment de honte chez Elster

Nous avons vu que, chez Freud, le « malaise » dans la civilisation vient d'un affadissement de la vie émotionnelle lié au renoncement aux satisfactions pulsionnelles les plus intenses. Cette position, nous venons de le voir, est mise en doute par l'approche culturaliste qui considère que le caractère et l'affectivité d'un individu obsédé par la satisfaction d'une pulsionnalité définie par une économie de la tension, de l'intensité et de la frénésie ne décrivent pas une condition humaine universelle et primitive, mais un type de psychologie individuelle résultant d'une certaine discipline culturelle des émotions. Pour les culturalistes, il n'y a donc pas de frustration liée à la socialisation *en soi* constituant une menace inévitable pour le bien-être de l'individu : il peut y avoir néanmoins une plus ou moins grande convergence entre les tendances affectives spontanées des individus et les coordonnées culturelles.

Dès lors, les culturalistes admettent que l'ordre socioculturel des normes et des valeurs peut exercer une certaine forme de violence sur l'individu lorsqu'il le contraint à adopter des attitudes qui vont à l'encontre de ses tendances affectives premières. Le timide peut être contraint de s'exposer, l'exubérant peut se voir intimer l'ordre de se retenir, le doux peut être forcé d'être agressif, le luxurieux peut être réduit à l'abstinence et celui qui goûte peu la sexualité peut être tenu d'afficher une certaine grivoiserie.

Le point commun existant entre les théories culturalistes et la théorie freudienne est la question de l'acceptabilité sociale de l'attitude morale adoptée par l'individu. Dès lors, l'idée d'une intériorisation par l'individu de la désapprobation morale d'autrui dont le jugement est standardisé par la configuration culturelle repose également sur une manifestation émotionnelle se présentant comme l'effectivité de l'évaluation sociale intériorisée. Ainsi, l'angoisse freudienne, quoique très marquée par le contexte de la théorie psychanalytique des pulsions, peut toutefois être réinvestie dans une théorie culturaliste de la normativité au moyen du concept de « honte » comme on l'observe chez Jon Elster (1998, 2008). Comme l'angoisse chez Freud, la honte est présentée par Elster comme une émotion très puissante liée à une évaluation morale négative de soi ainsi qu'à la crainte d'être rejeté par les autres.

Il est important de noter que l'angoisse décrite par Freud et la honte décrite par Elster ont deux points communs : d'une part la crainte d'être ostracisé par autrui, d'autre part le rapport à soi-même en tant qu'ego. Dans les deux cas, le rapport évaluatif à soi-même implique le détour par l'altérité des normes sociales morales. De ce point de vue, les théories de Freud et d'Elster sont très proches. On remarque donc que, si l'approche culturaliste développée en anthropologie dès les années 1930 tend à invalider l'idée selon laquelle le processus de socialisation reposerait toujours sur l'imposition d'un régime ascétique de gestion des pulsions, en revanche l'ajustement de l'individu à la normativité morale reste pensé à partir du mode de la sanction morale rendue effective par les états

émotionnels ressentis par les individus : « les émotions soutiennent et prêtent une efficacité causale aux normes sociales qu'autrement elles n'auraient pas eue¹⁴ ». La honte est une émotion reposant sur l'évaluation morale de soi en lien avec le regard désapprobateur d'autrui ; son pendant, la fierté, est en lien avec le regard approuvateur d'autrui. La crainte de la honte impliquant l'humiliation de soi et le sentiment du rejet de la part d'autrui se présente ainsi comme un mécanisme articulant émotion et évaluation morale.

III. INSTITUTION DE LA NORMATIVITÉ SOCIALE : LES MÉCANISMES D'ÉTABLISSEMENT DE LA NORMATIVITÉ SOCIALE SONT-ILS SEULEMENT MORAUX ?

A/ Radcliffe-Brown, Bateson et Goffman : les mécanismes cognitifs d'établissement de la normativité sociale structurale

L'anthropologie sociale et structurale développée par Radcliffe-Brown dès les années 1920 et 1930 permet d'envisager un autre mécanisme d'intériorisation et d'établissement de la normativité sociale dont les caractéristiques ne sont pas morales. L'intérêt de cette approche réside dans le fait qu'elle permet d'expliquer une part importante de ce que l'on nomme « normativité sociale » et qui n'appartient pas directement au champ de la moralité. En effet, les normes de politesse et tout ce que Goffman décrit comme des « rites d'interactions » ne renvoient pas directement au champ de la morale. Comme on le constate à la lecture de ses textes, les écarts et les maladroites commis par un acteur social peu compétent produisent plutôt l'embarras (*embarrassment*) que la honte. L'émotion générée est ici bien moins aiguë et le faux pas ne conduit pas à un ostracisme aussi radical que celui provoqué par les comportements moralement réprimés.

L'anthropologie sociale de Radcliffe-Brown repose sur l'idée que la socialité humaine doit être comprise à partir de l'idée de « structure », qui dans ses travaux se rapporte à un ensemble cohérent et stabilisé de formes interactionnelles que les individus adoptent selon la position qu'ils occupent dans la société, position que Radcliffe-Brown (1952) désigne par le terme *personnalité* utilisé non dans un sens psychologique, mais dans un sens social. La vie sociale se présente ainsi comme un jeu de rôle au sein duquel chacun, quelle que soit sa psychologie intime, doit jouer une partition relativement standardisée. Un individu traverse donc la structure sociale en changeant à plusieurs reprises de rôle, apprenant successivement à se comporter *en tant qu'*enfant, puis adolescent (de telle catégorie sociale), puis adulte (de telle catégorie sociale, faisant telles études, ayant tel métier, etc.) *en face de* telle ou telle personne. Comme l'explique Radcliffe-Brown, la normativité du jeu de rôle s'opère *dans l'interaction* avec un autrui qui occupe lui aussi une position structurale précise : l'interaction normale d'un enfant de huit ans n'est pas la même selon que l'individu avec lequel il interagit est un autre enfant du même âge, un grand frère, un parent, un enseignant, etc.

Comme l'a montré Bateson (1936), le contexte culturel tel que l'a défini Benedict influe sur les normes interactionnelles structurales en vigueur dans une société. Autrement dit, les normes de comportement entre un enfant et un parent ou entre un subalterne et son supérieur hiérarchique ne sont pas les mêmes dans les différentes sociétés. Néanmoins, comme le montre Goffman (1967), il existe toujours des procédures, des signes, des codes et des effets de ritualité permettant aux individus de clarifier la valeur sociale de leur interaction et de statuer sur leurs positions respectives dans la structure sociale. Lorsqu'un élève s'adresse respectueusement à un enseignant, il rappelle une règle structurale en même temps qu'il se reconnaît dans le rôle de l'élève et reconnaît à autrui le rôle du professeur. La structure sociale est ainsi établie par le rappel constant fait aux règles du jeu (dont elle est composée) au gré des interactions.

L'observation des rites et de codes sociaux montre que la normativité sociale est loin d'être strictement morale et n'implique pas toujours la reconnaissance ou la condamnation de la valeur morale de l'ego individuel. Contrairement à la honte décrite par Elster, l'embarras documenté par Goffman n'est pas lié à un jugement moral. Autrement dit, l'embarras sanctionne un faux pas qui atteste d'un manque de *compétence* de la part de l'acteur social, non de sa *méchanceté* morale. La maladresse ne conduit pas à l'infamie.

La mise en lumière de ce second cas de figure nous permet donc de montrer ici deux choses : en premier lieu, qu'il faut distinguer deux aspects de ce que l'on nomme la normativité sociale, à savoir la normativité sociale morale et la normativité sociale structurale (c'est-à-dire l'ensemble des codifications interactionnelles ayant cours dans une société donnée) ; en second lieu, qu'il faut alors distinguer deux modalités émotionnelles différentes contribuant à l'établissement de la normativité sociale, à savoir la peur de la honte et la crainte de l'embarras. Dans le premier cas, l'individu cherche à éviter une condamnation morale impliquant une sanction émotionnelle brutale, tandis que dans le second cas l'individu veut passer aux yeux des autres comme un interactant social compétent et éviter une sanction émotionnelle bien plus légère.

Il est important de noter ainsi que l'établissement des aspects structuraux de la normativité sociale implique un rapport moins aigu à la valorisation de l'ego, car ce rapport n'est pas tant moral que cognitif. En effet, le maladroit ne risque pas d'être ostracisé comme un monstre moral, mais seulement d'être méprisé à partir d'une évaluation cognitive : il sera vu comme un importun et un « idiot » (le terme renvoyant étymologiquement à celui qui peine à se départir de ses usages locaux afin de maîtriser une gamme élargie et complexe de règles interactionnelles).

Faut-il alors en conclure que la normativité sociale structurale est beaucoup plus fragile que la normativité sociale morale, puisque le sentiment d'embarras est nettement plus faible que celui de honte ? Un premier élément de réponse à cette question repose sur l'observation du lien existant entre normativité morale et normativité structurale. Ce lien est souligné par Goffman (Goffman, 1967) qui

note qu'il existe des aspects moraux au « *facework* », c'est-à-dire à l'effort interactionnel consenti par les membres d'une société pour garantir la reconnaissance mutuelle de leur personnalité sociale. Cela signifie que les règles morales s'appliquent différemment en fonction de la personnalité sociale de l'individu concerné : par exemple s'il est peu gênant pour une star du cinéma et ou de la musique d'afficher une vie de plaisir et de sexualité relativement débridée dans une société occidentale contemporaine, il est beaucoup moins acceptable socialement de le faire pour une personnalité politique. Ainsi, évoquer un tel mode de vie concernant une personnalité publique pourra être perçu comme un clin d'œil amusé dans le premier cas, mais comme une attaque brutale dans le second. Par conséquent, si les aspects moraux et structuraux semblent devoir être distingués par l'analyse intellectuelle, il importe de noter qu'ils sont souvent entremêlés dans la pratique quotidienne, puisque les différentes positions structurales impliquent des variations *ad hoc* dans l'application des normes morales d'une société.

Faut-il alors considérer que la normativité sociale structurale est principalement établie de façon indirecte par l'effet de la normativité sociale morale soutenue par le sentiment de honte ?

B/ Bateson et Hilgard et Marquis : apprentissage, anxiété et structure de l'environnement cognitif

Nous voulons ici montrer qu'il existe un mécanisme d'établissement de la normativité sociale structurale qui lui est propre et qui est bien distinct des enjeux moraux de la honte et de la fierté que peut éprouver le moi ayant intériorisé les valorisations et les condamnations morales de sa culture. Ce second mécanisme, loin de relever des théories du sentiment moral, relève plutôt des théories de l'apprentissage. Le premier à travailler sur cette option théorique est Gregory Bateson qui, dès la fin des années 1940, s'intéresse aux travaux des psychologues et des behavioristes afin d'approfondir la réflexion sociale et structurale inaugurée par Radcliffe-Brown en anthropologie. En décrivant la société comme un corpus systématique de normes et de rôles stéréotypés ayant cours dans les groupes humains, l'approche structurale se trouve confrontée au problème de l'explication de l'adhésion des individus aux règles du jeu social. En montrant qu'un individu s'intègre à une société en s'efforçant de devenir un acteur social compétent, Bateson (1936, 1972) et Goffman (1967) mettent le doigt sur un phénomène décisif pour comprendre la socialisation des êtres humains, c'est-à-dire le fait que l'individu humain cherche spontanément à savoir *comment il doit se conduire* et se trouve plongé dès sa prime enfance au sein d'interactions sociales dont les règles, les formes et les valeurs implicites sont à découvrir.

Le ressort démonstratif d'une telle position repose sur l'idée selon laquelle tout individu humain adhère *a priori* au jeu social et qu'il a intérêt à le faire. Bateson et les psychologues de l'apprentissage que sont Hilgard et Marquis mettent alors en lumière un phénomène de socialisation bien différent de celui reposant sur la gestion de la frustration et l'intériorisation des normes sociales d'évaluation morale par l'individu. Pour Bateson, qui s'intéresse à la question de l'inté-

gration des règles de comportement structurales, la socialisation des individus doit être – sur ce point précis – assimilée à l'apprentissage d'un ensemble de jeux codifiés. Il ne s'agit donc pas ici de l'effort plus ou moins douloureux d'ajustement des aspirations spontanées de l'individu avec le contexte moral de la culture en question, mais des *capacités cognitives* de l'individu qui maîtrise avec plus ou moins d'acuité, d'aisance et de subtilité les codes de la société à laquelle il appartient¹⁵.

Bateson (1972) montre que l'on peut établir une connexion particulière entre l'aisance cognitive et un type spécifique de tonalité affective. Certaines analyses du comportement animal qu'il trouve chez Hilgard et Marquis (1940) lui permettent notamment de cerner deux sentiments particuliers – confiance et anxiété – qui sont des sentiments manifestant un rapport abstrait au monde et à sa structure. Bateson note ainsi qu'il existe une forme de confiance qui se présente comme un sentiment de sécurité résultant de la maîtrise intellectuelle de la logique du monde, c'est-à-dire d'une forme de familiarité avec ses règles et sa sémiotique¹⁶. À ce sentiment s'oppose l'anxiété qui sourd du doute et du constat d'impuissance d'un être capable de percevoir son incapacité à saisir les règles du jeu qui gouvernent la situation interactionnelle à laquelle il est confronté¹⁷.

Bateson s'appuie sur les travaux que Hilgard et Marquis ont consacrés à l'apprentissage, dans le champ de la psychologie expérimentale, pour approfondir cette question. Les sentiments de confiance et d'anxiété mis en relief par Bateson, Hilgard et Marquis ne sont pas des sentiments moraux, mais des affects qui procèdent de l'évaluation par un organisme de sa *capacité* cognitive à saisir un environnement ou, au contraire, du constat de sa désorientation. L'anxiété, en tant que sentiment général portant sur la façon dont l'organisme évalue son degré de compréhension d'un environnement, trahit une enquête à un niveau abstrait, c'est-à-dire au niveau relatif à la structuration de son environnement. L'anxiété, qui induit une injonction à réviser les règles comportementales admises, est considérée comme un élément majeur de l'apprentissage par Hilgard et Marquis dans *Conditioning and learning* (1940) : « L'anxiété fournit une base motivationnelle à l'adaptation dans la mesure où elle se présente comme une expérience déplaisante : ainsi, le soulagement de l'anxiété est considéré comme un objectif majeur¹⁸. » Les psychologues montrent que le processus de primo-socialisation peut générer une anxiété aiguë chez le jeune enfant humain en raison des difficultés qu'il éprouve à s'adapter à un environnement social et à sa logique. Inversement, la compréhension de l'environnement, c'est-à-dire des règles qui le structurent, permet de transformer ce sentiment angoissant en une rassurante sensation de maîtrise.

Hilgard et Marquis, afin d'approfondir leurs travaux sur l'apprentissage et la formation des habitudes sociales chez les humains, font un détour par ce que l'on nomme désormais l'éthologie et s'intéressent aux problèmes cognitifs et comportementaux rencontrés par d'autres mammifères, et notamment, à la suite des travaux de Pavlov, les chiens. Leurs études permettent de mettre en lumière la complexité des phénomènes d'adaptation cognitive auxquels les canidés

peuvent être confrontés. Hilgard et Marquis étudient ainsi les cas de névrose expérimentale, c'est-à-dire d'anxiété extrême liée à l'impuissance ressentie par un animal face à une situation cognitivement inextricable. Les psychologues se réfèrent à l'expérience menée par Shenger-Krestovnikova dans les laboratoires de Pavlov. Au cours de cette expérimentation, un chien – qui n'est pas affamé – se voit donner de la nourriture s'il est capable de faire la distinction entre un cercle et une ellipse. Progressivement, des figures de plus en plus ressemblantes sont présentées à l'animal de sorte qu'il ne parvienne plus à les discriminer clairement. Parvenu à ce moment de l'expérience, le chien devient nerveux, agressif et anxieux, pourtant il est important de rappeler qu'il ne souffre pas de la faim, mais simplement du fait d'être désorienté et cognitivement perdu¹⁹. Cette expérience montre un chien paniqué, car incapable d'affronter intellectuellement ce type de situation : il ne parvient pas à comprendre la règle qui structure cette situation et à statuer sur le type de réponse attendue. Il abandonne, perdant la capacité d'apprendre tandis qu'une effroyable anxiété fait place à la confiance. Le chien manifeste alors des signes évidents de stress ainsi que le rapportent Hilgard et Marquis²⁰. Le chien est acculé : apprendre n'apparaît plus comme une option possible. Seules l'angoisse et la colère demeurent. Pourtant il ne meurt pas de faim, mais se confronte à un monde dont les règles et les jeux sont au-delà de ses capacités intellectuelles. Le sentiment d'incapacité cognitive qui s'ensuit constitue *en soi* une punition, puisqu'il génère un stress puissant.

C/ Apprentissage et bien-être

Dès lors, si le doute qui vient troubler la quiétude d'une autosatisfaction rassurante peut pousser l'individu à enquêter sur les raisons de ses échecs (Peirce, 1878) et si les conflits cognitifs peuvent être féconds pour le progrès intellectuel de l'individu (Doise et Mugny, 1981, 1997), il apparaît que la dimension cognitive positive de l'échec implique que celui-ci puisse être surmonté par l'individu. En revanche, une difficulté perçue comme insurmontable à un moment donné produit une angoisse aiguë susceptible d'hypothéquer les capacités d'apprentissage de l'individu. Le cas de névrose expérimentale décrit par Hilgard et Marquis montre qu'un sujet profondément destabilisé quant à ses compétences est submergé d'une anxiété qui cause une diminution, parfois drastique, de ses capacités cognitives. La répétition d'échecs perçus comme insurmontables plonge l'animal dans un état d'anxiété croissant qui le conduit à interpréter l'expérience comme une violence à son égard générant ponctuellement une réaction émotionnelle de colère puis une apathie angoissée : le chien cesse de réagir de façon adéquate face à de nouvelles situations d'apprentissage. Il refuse désormais de jouer à toute forme de jeu : l'agressivité et l'apathie expriment ce refus.

Il n'est pas question dans ce cadre théorique de répression de tendances pulsionnelles ainsi que d'évaluation morale de la valeur de l'individu : frustration, mal-être et angoisse sourdent ici de l'autoévaluation d'un sujet incapable de trouver la quiétude dans une confiance suffisante en ses capacités cognitives. Le pendant

objectif de cette évaluation négative est l'effritement du sentiment de familiarité ressenti par l'individu face aux situations ludiques ou aux situations sociales dont il fait l'expérience.

Appliquées aux questions d'apprentissage humain et de primo-socialisation, ces analyses de Hilgard et Marquis, que reprend Bateson (1972), tendent à montrer qu'un enfant humain sera soucieux de se plier aux exigences sociales dans la mesure où il désire maîtriser cognitivement les situations sociales auxquelles il est confronté afin d'éprouver le sentiment rassurant de confiance que l'on ressent dans un environnement familial. Le désir de se familiariser avec le milieu social, d'en maîtriser les règles et les codes apparaît ainsi comme un ressort extrêmement puissant d'institution des normes sociales structurales. Comprendre comment il faut agir afin de devenir un acteur social compétent est un souci qui précède très largement l'attitude réflexive ultérieure que peuvent adopter certains rares individus se posant la question de la légitimité et de l'acceptabilité de ces normes. Si l'on prend l'exemple d'un enfant qui entre dans un nouvel établissement scolaire, on remarque qu'il sera inévitablement stressé et que son premier souci sera donc de comprendre quelles sont les règles officielles et non officielles qui structurent cet environnement encore non familial. Son intégration repose ainsi sur sa capacité à saisir les normes et les codifications locales, ce qui est en soi rassurant, tout en permettant également de démontrer ses compétences sociales aux autres et donc d'être admis dans la communauté.

De surcroît, on a noté (Saarni, 2022) que l'apaisement produit chez l'enfant par la perception d'un environnement familial et prévisible aide le sujet dans l'apprentissage de la gestion de ses émotions, notamment de sa colère²¹. Ainsi, la clarté cognitive de l'environnement structural permet le déploiement d'une « affectivité de basse intensité²² » (Quéré, 2021) propice non seulement à l'apprentissage intellectuel, mais également à l'apprentissage de la gestion émotionnelle. On note donc un nouveau lien existant entre normativité sociale structurale et normativité sociale morale : la standardisation morale des affects individuels en fonction des réquisits axiologiques moraux de la culture d'une société semble pouvoir être facilitée par la disponibilité cognitive et émotionnelle que permet la reconnaissance d'un environnement structural familial et maîtrisé chez l'enfant.

D/ Quéré lecteur de Garfinkel : maîtrise des conditions socialement structurées d'interaction et apaisement émotionnel

L'anxiété qu'accompagne la perte des repères sémiotiques du jeu social structural est décrite par Louis Quéré (2021) qui reprend l'exemple des « *breaching experiments* » de Garfinkel. Quéré rappelle que Garfinkel cherchait à « faire apparaître ces “conditions socialement structurées” de production des émotions sociales » en demandant à ses étudiants « de faire dans leurs interactions avec d'autres personnes “ce qui ne se fait pas” : se comporter chez soi comme un étranger, en respectant une formalité qui n'est pas d'usage » (Quéré, 2021, p. 139). Quéré rappelle que ces expériences suscitaient à l'avance « beaucoup

d'appréhension et d'anxiété chez les étudiants » et il rapporte la sidération et la désorientation que décrivent ceux qui s'y sont confrontés. Il explique ainsi :

La déception des attentes et la contradiction des orientations constitutives de l'engagement dans une situation sont source d'émotions ; elles provoquent notamment des surprises désagréables, voire de la sidération car elles génèrent de l'indétermination et de l'incertitude. Mais l'expérimentation de Garfinkel donne aussi à penser que la dialectique de la confiance et de la méfiance joue un rôle important dans les conditions sociales de production des émotions. En un sens, il n'y a là rien de très étonnant si l'on considère que la confiance permet de réduire l'incertitude inhérente aux comportements futurs d'autrui, et que la substitution de la méfiance à la confiance démultiplie les incertitudes, bloque les modes d'action et les modes de coordination ordonnés et établis, et engendre de l'embarras et de l'impuissance dans la gestion des interactions et des émotions qui vont avec. (Quéré, 2021, p. 143)

Ces observations de Quéré effectuées à partir des « *breaching experiments* » de Garfinkel tendent à corroborer les conclusions de Hilgard et Marquis ainsi que celles de Bateson à propos des névroses expérimentales. Dans les deux cas, l'anxiété est générée par un procédé visant à décevoir les attentes normales et normatives d'un sujet pris dans une situation sociale *a priori* identifiée. Dans les deux cas, il est montré que la rupture avec les conditions socialement structurées de la situation (jeu bien établi en laboratoire dans le cas de l'expérimentation menée avec le chien ; forme conventionnelle et ordinaire d'une interaction sociale dans le cas des « *breaching experiments* ») provoque ce que les auteurs nomment une « anxiété » qui n'est pas seulement une émotion négative, mais qui se présente comme un bouleversement dans l'économie émotionnelle de sujet : celui-ci est alors submergé par une affectivité intense, négative et qu'il ne maîtrise pas. L'anxiété produit une fébrilité ainsi qu'une désorientation paniquée qui induisent l'apparition de la colère, de l'indignation, de l'embarras, voire d'une forme de fuite et d'apathie, qui sanctionnent un jugement de l'acteur sur sa compétence sociale. Pour les autres, celui-ci n'est pas mauvais moralement, mais inapte socialement, ce qui peut parfois relever de ce que l'on qualifie communément par le terme *folie*.

Ces observations permettent de mettre au jour un mécanisme puissant d'imposition de l'ordre social et des rituels structuraux de la vie quotidienne. L'apprentissage de ces derniers est une condition nécessaire à l'intégration de l'individu dans un univers familier et rassurant qui lui permet de déployer une affectivité de basse intensité propice d'une part à l'apprentissage cognitif et d'autre part à l'apprentissage de la gestion des émotions. L'apprentissage de la normativité sociale structurale qui conduit à son respect spontané de la part des individus est ainsi motivé par un désir aigu de clarté cognitive et de familiarité qui concerne tous les êtres humains. L'intérêt des individus pour une telle situation de confort cognitif et émotionnel est majeur puisqu'il est une condition décisive de ce que l'on pourrait décrire comme la disponibilité mentale nécessaire

aux efforts adaptatifs, qu'ils soient cognitifs ou émotionnels. Cette disponibilité mentale conférée par la familiarité d'un environnement sémiotique et normatif se présente donc comme le corrélat du bien-être entendu comme tonalité affective de basse intensité susceptible d'être décrite comme une forme générale de confiance.

CONCLUSION

Il apparaît, à l'issue de cette analyse, qu'il est important pour la réflexion philosophique portant sur le bien-être de l'individu en société de tenir compte de la distinction entre normativité sociale morale et normativité sociale structurale.

En ce qui concerne la normativité sociale morale, nous avons vu que l'on devait, sans totalement abandonner le modèle freudien de la répression des satisfactions pulsionnelles par l'angoisse, prendre en compte les critiques à son égard issues du culturalisme. Les effets de contrainte produits par les normes morales sur les individus sont, certes, susceptibles de varier en fonction des tendances innées des individus et du type de configuration culturelle au sein duquel il est intégré. Néanmoins, l'idée qu'une punition émotionnelle résultant de l'intériorisation du jugement d'autrui – angoisse ou honte – est une source de motivation puissante au respect des valeurs morales imposées par la culture reste convaincante, comme le montrent les travaux d'Elster. On peut alors penser le degré de bien-être de l'individu en fonction de la convergence ou de la divergence existant entre les tendances caractérielles innées de l'individu et les exigences morales standardisées de la culture.

En ce qui concerne la normativité sociale structurale, nous avons vu qu'il fallait éviter de réfléchir en termes de sanction morale, mais privilégier une approche insistant sur le phénomène de l'apprentissage. Il est alors apparu que la clarté cognitive subjectivement perçue tendant à faire de l'environnement social, avec ses règles et ses signes, un milieu familial et rassurant est un ressort majeur de la socialisation. Ce sentiment de confiance est apparu comme une tonalité affective de basse intensité favorisant la cognition et la gestion des émotions. Réciproquement, l'expérience de la désorientation est apparue comme générant une tonalité affective de forte intensité caractérisée par une sorte de fébrilité et de panique que l'on peut nommer anxiété. Nous avons vu que cette anxiété tend à inhiber les processus cognitifs d'apprentissage et à déranger les processus affectifs de gestion des émotions. La question de la normativité sociale structurale permet donc de mettre en lumière un lien fort existant chez l'être humain entre la clarté cognitive et la gestion des émotions qui permet de comprendre pourquoi la maîtrise cognitive d'un environnement normatif tenu pour familier est une source majeure de bien-être. Ce lien établi entre respect des rites sociaux d'interaction et bien-être ouvre également la voie à une meilleure compréhension des mécanismes d'institution des normes sociales. Le sentiment de confiance résultant de la maîtrise des règles du jeu social apparaît comme un puissant élément de motivation induisant les individus à respecter inconditionnellement les normes comportementales en cours dans leur communauté.

L'approche structurale issue de la sociologie et de l'anthropologie sociale nous semble donc apporter un complément important à l'approche morale que l'on retrouve en psychanalyse ou dans le cadre des théories des sentiments moraux. L'articulation entre les théories morales de la valorisation et les théories de l'apprentissage nous apparaît comme une voie féconde d'approfondissement des réflexions sur la socialisation et l'agir social. Le lien que l'on peut établir entre familiarité environnementale, clarté cognitive et gestion des émotions permet notamment d'articuler la question structurale de l'adaptation aux institutions et la question morale de la standardisation culturelle des émotions et des valeurs. Le point central sur lequel il convient d'insister d'ores et déjà est qu'il ne faut pas seulement prendre en compte la sanction morale dans la réflexion sur l'institution des normes sociales, mais aussi l'intérêt directement cognitif qui incite le sujet à désirer être un acteur social compétent. Dès lors, deux éléments distincts semblent devoir entrer dans la description du bien-être : d'une part, le sentiment de valorisation résultant d'une évaluation morale dont les standards sont intériorisés par le sujet et ceux qui l'entourent ; d'autre part, l'établissement d'une affectivité de basse intensité caractérisant un lien de familiarité entre le sujet et son environnement. Cette forme de confiance²³ permet une disponibilité mentale accrue chez l'individu qui l'installe dans un cercle vertueux de réussite tant sous le rapport de la cognition que sous celui de la gestion de ses émotions.

NOTES

- ¹ Le modèle freudien a connu un certain succès en anthropologie durant la première moitié du XX^e siècle, notamment dans l'œuvre de Bronislaw Malinowski qui, en 1927, publie *Sex and Repression in Savage Society* (Malinowski, 1927).
- ² Freud, 2016 [1929], p. 45-46 : « Nous avons signalé plus haut ce fait d'expérience que l'amour sexuel (génital) procure à l'être humain les plus fortes satisfactions de son existence et constitue pour lui à vrai dire le prototype de tout bonheur ; et nous avons dit que de là à rechercher également le bonheur de la vie dans le domaine des relations sexuelles et à placer l'érotique génitale au centre de cette vie, il aurait dû n'y avoir qu'un pas. »
- ³ Freud, 2016 [1929], p. 19 : « Ce qu'on nomme bonheur, au sens le plus strict, résulte d'une satisfaction plutôt soudaine de besoins ayant atteint une haute tension, et n'est possible de par sa nature que sous forme de phénomène épisodique. [...] nous sommes ainsi faits que seul le contraste est capable de nous dispenser une jouissance intense, alors que l'état lui-même ne nous en procure que très peu. »
- ⁴ Freud, 2016 [1929], p. 22 : « Ceci ne signifie nullement qu'on ait renoncé à toute satisfaction, mais bien qu'on s'est assuré une certaine garantie contre la souffrance en ceci que l'insatisfaction des instincts (*Trieben*) tenus en bride n'est plus ressentie aussi douloureusement que celle des instincts (*Trieben*) non inhibés. En revanche, il s'ensuit une diminution indéni- niable des possibilités de jouissances. La joie de satisfaire un instinct resté sauvage, non domestiqué par le Moi, est incomparablement plus intense que celle d'assouvir un instinct dompté. Le caractère irrésistible des impulsions perverses, et peut-être l'attrait du fruit défendu en général, trouvent là leur explication économique. »
- ⁵ Freud, 2016 [1929], 30-31 : « On découvrit alors que l'homme devient névrosé parce qu'il ne peut supporter le degré de renoncement exigé par la société au nom de son idéal culturel, et l'on en conclut qu'abolir ou diminuer notablement ces exigences signifierait un retour à des possibilités de bonheur. »
- ⁶ Freud, 2016 [1929], 70-71 : « Un grand changement intervient dès le moment où l'autorité est intériorisée, en vertu de l'instauration d'un Surmoi. Alors les phénomènes de conscience (morale) se trouvent élevés à un autre niveau, et l'on ne devrait parler de conscience et de sentiment de culpabilité qu'une fois ce changement opéré. Dès lors l'angoisse d'être découvert tombe aussi, et la différence entre faire le mal et vouloir le mal s'efface totalement, car rien ne peut rester caché au Surmoi, pas même des pensées. Toutefois la gravité réelle de la situation a disparu du fait que l'autorité nouvelle, le Surmoi, n'a aucune raison, croyons-nous, de maltraiter le Moi auquel il est intimement lié. Mais l'influence de sa genèse dont le mode permet au passé, et au dépassé, de survivre en lui se manifeste en ceci qu'au fond tout est resté comme avant, dans l'état primitif le Surmoi tourmente le Moi pécheur au moyen des mêmes sensations d'angoisse et guette les occasions de le faire punir par le monde extérieur. »
- ⁷ Freud, 2016 [1929], 72 : « Nous connaissons ainsi deux origines au sentiment de culpabilité : l'une est l'angoisse devant l'autorité, l'autre, postérieure, est l'angoisse devant le Surmoi. La première contraint l'homme à renoncer à satisfaire ses pulsions. La seconde, étant donné l'impossibilité de cacher au Surmoi la persistance des désirs défendus, pousse en outre le sujet à se punir. »
- ⁸ Freud, 2016 [1929], 71-72 : « Nous discernons maintenant le rapport existant entre le "renoncement aux pulsions" et le sentiment de culpabilité. A l'origine, le renoncement est bien la conséquence de l'angoisse inspirée par l'autorité externe ; on renonce à des satisfactions pour ne pas perdre son amour. »
- ⁹ Bateson, 1934 : « It is, I think, time that psychologists and anthropologists carefully examined the oft-repeated statement that man is innately savage and born with a tendency to warfare. The statement is generally supported by saying that his ancestors were savage war-making creatures and that European man has inherited their aggressive instincts. These statements are implicit in Dr. William Brown's letter in *The Times* of December 11. He says, "the same tendencies [to self-assertion, aggression, &c.] are there, lying hidden but not entirely dormant, in times of peace", and goes on to argue that therefore we must prepare for war. »

- ¹⁰ Bateson, 1934 : « First let us consider the supporting facts. Is it certain that man's ancestors were particularly savage? If we study primitive man the position is perfectly plain. I have myself spent some years in New Guinea studying three primitive peoples, and the people of whom I know most were head-hunters. The position is simply this: Some primitive communities (for example the Arapesh studied by Dr. Fortune in New Guinea, the Sulka whom I have myself studied, the Zuni in Mexico, and many others) are almost entirely non-aggressive. Others – and here we find the majority of primitive peoples – have fitted this business of fighting and aggression into their culture; they have built up round it a series of rules and conventions with the result that in many cases the fighting is more closely allied to a sport or spectacle than to warfare as we know it. Lastly, there is a very small minority of primitive peoples who have developed something resembling the warfare of Europe. As examples of this last category we may take the Zulus under Caka and the Lae Womba of New Guinea, who under the influence of some sort of religious revival went mad on war in the early years of the twentieth century. From facts of this sort it is evident that we cannot ascribe European warfare directly to man's innate aggressiveness. There is clearly some cultural factor at work which effectively determines whether a community shall habitually fight and what sort of fighting it shall have. »
- ¹¹ Benedict, 1960 [1934], p. 218 : « In reality, society and the individual are not antagonists. »
- ¹² Benedict, 1960 [1934], p. 219 : « The man in the street still thinks in terms of a necessary antagonism between society and the individual. In large measure this is because in our civilization the regulative activities of society are singled out, and we tend to identify society with the restrictions the law imposes upon us. The law lays down the number of miles per hour that I may drive an automobile. »
- ¹³ Quéré, 2021, p. 224 : « The tribes we have described have all of them their non-participating “abnormal” individuals. »
- ¹⁴ Elster, 2015 : « Je ne pense pas que les normes sociales en général surgissent pour “contre-carrer” les tendances spontanées. Il est d'ailleurs difficile de savoir comment et pourquoi elles apparaissent et parfois disparaissent. Mais on peut en démontrer des modes d'opération. En ce qui concerne la vengeance, on peut par exemple noter que, dans certaines sociétés tout au moins, attendre est un signe de motivation prudente qui peut être assimilé à la peur : quelqu'un qui attend pour se venger, c'est aussi quelqu'un qui pourrait avoir peur d'être lui-même tué ; donc il y a aussi des normes sociales qui vont dans le sens d'une injonction à la violence immédiate. La tendance à se venger immédiatement peut ainsi provenir à la fois de l'émotion et de l'injonction à sauver son honneur. Mais dans le second cas, il s'agit aussi d'émotion, puisqu'on ne veut pas s'exposer au mépris ; ce sont deux mécanismes émotionnels distincts qui peuvent conduire au même résultat : l'action immédiate. Distinguer nettement normes sociales et émotions comme types de motivation est donc très délicat puisque ce sont les émotions qui sont les moteurs des normes sociales. L'efficacité causale des normes sociales vient des émotions de mépris dans l'observateur et de honte dans l'agent. Les émotions soutiennent et prêtent une efficacité causale aux normes sociales qu'autrement elles n'auraient pas eue. »
- ¹⁵ Dans *Naven* (1936) Bateson décrit ainsi une discussion entre des étudiants de Cambridge qui pratiquent une forme d'humour à froid susceptible d'être déstabilisant pour un non initié. Bateson montre que l'attitude sérieuse et grave des interlocuteurs est une feinte qui renforce l'effet drolatique de leurs propos absurdes. Un nouveau venu pourrait facilement être égaré par cette attitude et être ainsi moqué pour son incompetence.
- ¹⁶ Pour Bateson, tout animal manifeste un comportement car son environnement est empli de signes, de valeurs et d'indications de toutes sortes qui déclenchent des séquences comportementales de différents types. Selon lui, se comporter consiste à agir en fonction de signes dont la signification est dépendante de l'anticipation de séquences récurrentes. Si la sémiotique sociale humaine est plus riche, plus complexe et plus subtile que celle des autres animaux, elle repose néanmoins sur un type de compétences cognitives partagé par d'autres espèces animales.

- ¹⁷ On notera la similarité entre le doublet confiance / anxiété présenté par Bateson et l'opposition certitude / doute que l'on trouve chez Peirce.
- ¹⁸ Hilgard et Marquis, 1940, p. 439, nous traduisons : « Anxiété : L'anxiété est importante [...] dans la majorité des sens selon lesquels la théorie de l'apprentissage tend à mobiliser ce concept. L'anxiété fournit une base motivationnelle à l'adaptation puisqu'elles se présente comme une expérience déplaisante : ainsi, le soulagement de l'anxiété est considéré comme un objectif majeur. D'ailleurs, l'anxiété a été vue comme fournissant les signaux (stimuli) déclenchant les différents mécanismes adaptatifs. »
- ¹⁹ Hilgard et Marquis, 1940, p. 440, nous traduisons : « Névroses expérimentales : L'une des premières approches des problèmes de personnalité en laboratoire a consisté dans l'étude des *névroses expérimentales*, terme qui, bien que souvent critiqué, continue d'être utilisé (Liddell, 1956). Ce champ d'application des procédures et des théories du conditionnement a été ouvert par les observations expérimentales de Shenger-Krestovnikova dans les laboratoires de Pavlov en 1914 (Pavlov, 1927). Un chien fut entraîné à saliver quand un cercle lumineux était projeté sur un écran. Après que la réponse conditionnée eut été bien établie, on chercha à obtenir une discrimination entre un cercle et une ellipse dont les rayons admettaient un ratio de 2/1. La discrimination fut obtenue assez rapidement. La forme de l'ellipse fut ensuite progressivement modifiée jusqu'à se rapprocher fortement d'un cercle. Durant ces manipulations l'entraînement à la discrimination continua : le cercle donnait lieu à un renforcement positif et non l'ellipse. Enfin, une ellipse dont les rayons admettaient un ratio de 9/8 fut atteinte. À partir de ce moment la discrimination ne cessa pas seulement de progresser mais, durant les trois semaines d'entraînement qui suivirent, elle empira. »
- ²⁰ Hilgard et Marquis, 1940, p. 440, nous traduisons : « À partir de ce moment l'intégralité du comportement de l'animal se mit à changer. Le chien autrefois calme se mit à gémir dans son box, se contorsionnant pour s'échapper, déchirant avec ses crocs le harnais produisant les stimulations mécaniques sur la peau et mordant les tuyaux reliant la salle des animaux à celle des observateurs, ce qui n'avait jamais eu lieu précédemment. Lorsqu'il était amené dans la salle d'expérimentation le chien aboyait désormais avec violence, ce qui contrastait avec sa tranquillité initiale. Bref, il présentait les symptômes d'une condition que l'on nommerait névrose chez l'être humain. Parmi les symptômes additionnels rapportés par les examinateurs on citera (1) des signes d'anxiété tels que les gémissements et les tremblements, (2) un effondrement de la précision de la réponse conditionnée avec une latence de réponse extrêmement variable, (3) le refus du chien de manger dans le dispositif expérimental et dans la salle, (Dworkin, 1939) et (4) des signes d'"inhibition" forte tels que bâillements, somnolence et sommeil (Muncie et Gantt, 1938). »
- ²¹ Saarni, 2011, p. 7 : « Un enfant qui voit le monde comme un endroit imprévisible, non-réceptif, et/ou hostile doit dépenser beaucoup d'énergie à maîtriser ses réactions affectives. »
- ²² Quéré, 2021, p. 143 : « La certitude complète est peu propice à l'émotion. C'est sans doute pourquoi le genre d'assurance qui sous-tend le comportement immédiat est émotionnellement plat : il exclut l'inquiétude, le souci, le doute et le questionnement. Il génère dans les interactions une tonalité affective que l'on pourrait dire "au ralenti", de "basse intensité" ou à "faible énergie" : la quiétude de la familiarité, ou celle qui est liée à l'opérativité des habitudes. »
- ²³ Il pourrait être utile de prolonger cette réflexion en y intégrant l'étude des travaux de Niklas Luhmann sur la confiance (Luhmann, 2006 [1968]).

BIBLIOGRAPHIE

Bateson, Gregory, « Psychology and the War », *The Times*, 13 décembre 1934.

Bateson, Gregory, *Naven. A Survey of the Problems suggested by a Composite Picture of the Culture a New-Guinea Tribe drawn from Three Points of View*, Cambridge, Cambridge University Press, 1936.

Bateson, Gregory, *Steps to an Ecology of Mind*, New York, Ballantine Books, 1972.

Benedict, Ruth, *Patterns of Culture*, New York, Mentor Books, 1960 [1934].

Doise, Willem et Gabriel Mugny, *Le développement social de l'intelligence*, Paris, Inter-Editions, 1981.

———, *Psychologie sociale et développement cognitif*, Paris, Armand Colin, 1997.

Dworkin, Simon, « Conditioning Neuroses in Dog and Cat », *Psychosomatic Medicine*, vol. 1, n° 3, 1939, p. 388-390.

Elster, Jon, *Alchemies of the Mind. Rationality and the Emotions*, Cambridge, Cambridge University Press, 1998.

———, *Strong Feelings. Emotions, Addiction and Human Behavior*, Bloomsbury Academic, 2008.

———, « L'influence négative des émotions sur la cognition » in *Terrains/Théories*, n° 2, 2015. <https://doi.org/10.4000/teth.255>

Freud, Sigmund, *Au-delà du principe de plaisir*, Paris, Petite bibliothèque Payot, 1981 [1920].

———, *Malaise dans la civilisation*, « Les classiques des sciences sociales », Chicoutimi, 2016 [1929]. http://classiques.uqac.ca/classiques/freud_sigmund_2/malaise_civilisation/malaise_civilisation.pdf

Garfinkel, Harold, *Studies in Ethnomethodology*, Englewood Cliffs, NJ, Prentice-Hall, 1967.

Goffman, Erving, *Interaction Ritual. Essays on Face-to-Face Behavior*, New York, Anchor Books, 1967.

Hilgard Ernest R. et Donald-George Marquis, *Conditioning and Learning*, New York, Appleton-Century-Crofts, 1940.

Liddell, Howard S., *Emotional Hazards in Animals and Man*, Springfield, C.C. Thomas, 1956.

Luhmann, Niklas, *La Confiance. Un mécanisme de réduction de la complexité sociale*, Paris, Economica, 2006 [1968].

Malinowski, Bronislaw, *Sex and Repression in Savage Society*, Londres, Routledge & Kegan Paul, 1927.

Muncie, Wendel et W. Horsley Gantt, « Effect on Behavior of Inhibition of Different Forms of Excitation », *American Journal of Physiology*, vol. 123, n° 1, 1938, p. 152-158.

Pavlov, Ivan P., *Conditioned Reflexes*, London, Oxford University Press, 1927.

Peirce, Charles S., « How to Make Our Ideas Clear », *Popular Science Monthly*, n° 12, p. 286-302, 1878.

Quéré, Louis, *La fabrique des émotions*, Paris, PUF, 2021.

Radcliffe-Brown, Alfred, *Structure and Function in Primitive Society*, Glencoe, IL, The Free Press, 1952.

Saarni, Carolyn et Linda A. Camras, « Développement affectif chez l'enfant » in Tremblay, R.E., M. Boivin et R. DeV Peters (dir.), *Encyclopédie sur le développement des jeunes enfants*. <https://www.enfant-encyclopedie.com/emotions/selon-experts/developpement-affectif-chez-lenfant>, septembre 2022.