



Available online at  
**ScienceDirect**  
www.sciencedirect.com

Elsevier Masson France  
**EM|consulte**  
www.em-consulte.com



## PHILOSOPHICAL CONSIDERATIONS

# Kant, autonomy and bioethics



## *Kant, autonomie et bioéthique*

**L. Campbell (Lecturer in medical ethics)**

*Clinical Sciences Institute, National University of Ireland, University Road, H91 TK33 Galway, Ireland*

Received 24 March 2017; accepted 5 May 2017  
Available online 9 August 2017

### KEYWORDS

Autonomy;  
Bioethics;  
Kant;  
Philosophy

**Summary** The concept of autonomy has played a pivotal role in bioethics discourse since the 1970s. Yet, prior to the emergence of bioethics, autonomy had received scant mention in twentieth-century philosophy and was conspicuous by its absence from discussions of health-care. The term was not even mentioned in the 1967 edition of the *Encyclopedia of philosophy*. The emergence of bioethics in the early 1970s coincided with increased attention across the western world to civil and human rights; with the rise of this new discipline the liberal emphasis on individual rights was recast in terms of respect for patient autonomy. Although its legal appeal was based on the ease with which autonomy was operationalized in the doctrine of informed consent, the power of the concept of autonomy lay in what it symbolized: the right of an individual to resist coercion or compulsion in the context of a relationship of power. Most commentators in the field of bioethics are familiar with autonomy as one of the four principles of biomedical ethics laid down by Beauchamp and Childress in their canonical text, *The principles of biomedical ethics* (1979). ‘Principlism’ is a mid-level theoretical tool, which has had broad appeal in facilitating analysis of ethical dilemmas in biomedicine, grasped in the abstract as conflicts between two or more of the four principles. Yet the principle of autonomy, which has had such an extraordinary influence in contemporary bioethics bears only, passing resemblance to the concept of autonomy, which emerged in early modern philosophy. Although the bioethical redrawing of autonomy owes a large debt both to the philosophical tradition and to the social upheavals of the twentieth century, the relationship between contemporary interpretations of the concept of autonomy in bioethics and its historical origins is rarely examined. The purpose of this paper is to trace the evolution of the concept of autonomy from its emergence in modern moral theory to contemporary debates about its relevance for bioethical analysis. The roots of the principle of autonomy can be traced back to the political theory of ancient Greece. Originally used to describe the capacity of the Greek polis or city-state to govern itself, the concept of autonomy received its first modern expression – and its first application to the individual – in the moral theory of Immanuel Kant. For Kant, autonomy stood for

*E-mail address:* [louise.campbell@nuigalway.ie](mailto:louise.campbell@nuigalway.ie)

the ideal of free will: a human will be driven to action, not by appetite or desire, but by identification with a 'higher' or rational self. At the heart of Kant's ethics is the close association of moral action with human rationality; for Kant, autonomous action – action which is deliberately and self-consciously motivated by moral reasons – is the quintessential expression of human rationality. Although the moral universalism Kant sought to defend is no longer philosophically tenable, his insights about many of the core features of autonomous action remained influential until well into the twentieth century. This paper falls into four parts: in the first section I will explore the contextual factors which influenced the emergence of autonomy as a principle appropriate for bioethical analysis. From there, I will examine the hugely influential definition of autonomy put forward by Beauchamp and Childress in the *Principles of biomedical ethics* and trace the philosophical foundations of this concept. I will then provide a brief account of the concept of autonomy so central to Kant's moral theory and I will conclude by examining recent accounts of personal autonomy in contemporary philosophy with the aim of arriving at a richer understanding of autonomy, which can perhaps be of greater service to bioethics.

© 2017 Elsevier Masson SAS. All rights reserved.

## MOTS CLÉS

Autonomie ;  
Bioéthique ;  
Kant ;  
Philosophie

**Résumé** Le concept de l'autonomie a joué un rôle clé dans le discours sur la bioéthique depuis les années 1970. Pourtant, avant l'émergence de la bioéthique, l'autonomie a reçu peu d'attention dans la philosophie du 20<sup>e</sup> siècle et était absente dans les discussions en santé. Ce terme n'a même pas été abordé dans l'édition de 1967 de *l'Encyclopedia of philosophy*. L'émergence de la bioéthique au début des années 1970 coïncide avec l'accroissement, en Occident, de l'intérêt pour les droits civils et humains. Avec la montée de cette nouvelle discipline, l'emphase libérale sur les droits humains a été reformulée en termes de respect de l'autonomie du patient. Bien que l'intérêt légal soit axé sur la facilité avec laquelle l'autonomie a été mise en branle par la doctrine du consentement, le pouvoir du concept de l'autonomie demeure ce qu'il symbolise : le droit d'une personne de résister à la coercition ou à la compulsion dans le cadre d'une relation de pouvoir. La plupart des commentateurs dans le domaine de la bioéthique reconnaissent l'autonomie comme l'un des quatre principes de l'éthique biomédicale établis par Beauchamp et Childress dans leur texte canonique, sur les *Principles of biomedical ethics (1979)*. « Principlism » est un outil théorique de niveau intermédiaire qui a été largement utilisé pour faciliter l'analyse de dilemmes éthiques en biomédecine, définis de façon abstraite comme de conflits entre deux de ces quatre principes ou plus. Toutefois, le principe d'autonomie qui a eu une influence extraordinaire dans la bioéthique contemporaine ressemble peu au concept d'autonomie ayant émergé lors des balbutiements de la philosophie moderne. Bien que la reformulation bioéthique de l'autonomie soit en grande partie redevable envers la tradition de la philosophie et les soulèvements sociaux du 20<sup>e</sup> siècle, la relation entre les interprétations contemporaines du concept d'autonomie en bioéthique et ses origines historiques est rarement étudiée. Cet article a pour but de tracer l'évolution du concept d'autonomie, depuis son émergence dans la théorie morale moderne jusqu'aux débats contemporains quant à son importance dans l'analyse biomédicale. Les racines du principe d'autonomie s'étendent jusqu'à la théorie politique de la Grèce ancienne. Traditionnellement employé pour décrire la capacité de la « polis » grecque ou cité-État de se gouverner elle-même, le concept d'autonomie a premièrement été exprimé – et appliqué à la personne – dans la théorie morale d'Emmanuel Kant. Pour Kant, l'autonomie représentait l'idéal du libre arbitre : une volonté humaine menant à agir, non pas par appétit ou par désir, mais bien par identification avec un soi « supérieur » ou rationnel. Au cœur de l'éthique de Kant se trouve l'association étroite entre l'action morale et la rationalité humaine ; sa déontologie est un produit de son rationalisme. Pour Kant, l'action autonome – action qui est délibérément et consciemment motivée par des raisons morales – est l'expression quintessentielle de la rationalité humaine. Bien que l'universalisme moral que Kant tentait de défendre n'est maintenant plus philosophiquement défendable, ses réflexions sur les traits principaux de l'action autonome sont demeurées importantes jusqu'à très tard au 20<sup>e</sup> siècle. Cet article est divisé en trois parties. Premièrement, j'examine les premières versions du concept de l'autonomie telles qu'on les retrouve dans les théories philosophiques de Kant et de Rousseau. Je décrirai la relation entre ces théories et l'idée libérale première selon laquelle une personne est libre si ses actions sont d'une certaine façon les siennes. Ensuite, j'aborderai la transformation de cette interprétation initiale de l'autonomie dans la littérature bioéthique

ayant suivi la publication du *Belmont Report* pour ensuite étudier les différentes conceptions de l'autonomie comme liberté, authenticité et comme priorisation des 'intérêts critiques' du soi proposé par Tom Beauchamp, Martina Oshana, Gerald Dworkin et Harry Frankfurt respectivement. Finalement, je me pencherai sur la question à savoir si un récit de l'autonomie peut être récupéré de la philosophie à des fins d'analyse biomédicale, malgré la multitude d'interprétations de l'autonomie qui s'offre.

© 2017 Elsevier Masson SAS. Tous droits réservés.

## Introduction

The concept of autonomy has a 'tremendous resonance', not only in contemporary philosophical and political thought [1, p. 29], but also in bioethics. Recognition of the importance of autonomy has led to a seismic shift in the way healthcare is delivered to patients, irrevocably changing the face of medical decision-making over the course of the past five decades, and underpinning developments in health policy and regulation. A vital presupposition of the informed consent process, autonomy is viewed by some commentators as "by far the most significant value to have influenced the evolution of contemporary medical law" [2, p. 8]. The notion of autonomy allows us to accept moral and legal responsibility for our actions and to ascribe rights and responsibilities to others. The mandate to respect autonomy is explicitly acknowledged in Article 5 of the *Universal declaration on bioethics and human rights* (2005) and it is stated as one of the principles of the *Convention on the rights of persons with disabilities* (2006, Art. 3a)<sup>1</sup>. Political liberalism is "grounded in beliefs about the value of some variety of autonomy" [3, p. 12] and in many jurisdictions autonomy is acknowledged as one of the unenumerated constitutional rights of citizens.

Prior to the late eighteenth century, however, the term autonomy – meaning self-rule or self-government – was used to refer, not to individuals, but to independent city-states in the ancient world. Immanuel Kant revolutionized moral philosophy by using the term 'autonomy' to designate self-governing moral agents. As Kant's star waned in the late modern era, the concept of autonomy was consigned to the archives of philosophical history, failing even to secure a mention in the 1967 *Encyclopedia of philosophy* [4]. An "incredible surge of interest in biomedical ethics" in the 1970s [5, p. 55] led to the reappearance of the concept of autonomy, partly in response to social and political events.

Currently, the concept of autonomy is used in two principal contexts: as an analytic tool in bioethics and as an object of enquiry in contemporary moral and political philosophy. The concept of autonomy, which has become ubiquitous in bioethics discourse, is largely based on the pared-down interpretation put forward by Beauchamp and Childress as one of the so-called 'four principles' of biomedical ethics

[6]<sup>2</sup>. The concept of autonomy under scrutiny in academic philosophy is complex and ambiguous, anchored in discussions of freedom, determinism, moral responsibility and the nature of human agency. While the 'principlist' definition of autonomy is simplified and accessible, it lacks the philosophical substance to illuminate the deep structure of autonomous decision-making. By contrast, the discussion of autonomy in the academic literature is heavily theoretical and largely inaccessible to nonspecialists. In what follows I wish to ask whether, between the theoretical underdetermination of autonomy in the bioethics literature and the overdetermination of autonomy in the philosophical literature, there may be a middle ground which can still be of some service to bioethics.

This article has four sections: in the first section I will explore the contextual factors which influenced the emergence of autonomy as a principle appropriate for bioethical analysis. From there, I will examine the hugely influential definition of autonomy put forward by Beauchamp and Childress in the *Principles of biomedical ethics* and trace the philosophical foundations of this concept. I will then provide a brief account of the concept of autonomy so central to Kant's moral theory and I will conclude by examining recent accounts of personal autonomy in contemporary philosophy with the aim of arriving at a richer understanding of autonomy which can perhaps be of greater service to bioethics.

## Autonomy: the historical context

The emergence of the concept of autonomy as a focus for discussions of ethics in biomedicine was not an event which took place solely within the world of academic discourse. Several 'real-world' events contributed significantly to the perception of the importance of respect for autonomy as a foundational principle of bioethics in the US, including the campaigns mounted by civil rights activists clamouring for access to healthcare in the 1960s and 1970s and the scandal generated by the infamous Tuskegee Syphilis study in 1972. The reaction of the political and medical establishments to

<sup>1</sup> The concept of autonomy received no mention in the European declaration on human rights and fundamental freedoms of 1948.

<sup>2</sup> The term 'principlism' is frequently used in the literature to denote Beauchamp and Childress' "identification and elaboration of four fundamental moral principles: viz., autonomy, beneficence, non-maleficence (more commonly known as the 'harm principle'), and justice". See Arras, John, "Theory and bioethics; The Stanford Encyclopedia of Philosophy (Winter 2016 Ed.), Edward N. Z. (ed.), URL: <http://plato.stanford.edu/archives/win2016/entries/theory-bioethics>.

these events led to a shift away from social and medical paternalism towards an increasing acknowledgement of the role played by patients and research participants in determining what doctors and investigators should be allowed to do to them<sup>3</sup>.

### Patient rights

The late 1960s and early 1970s witnessed an unprecedented degree of social unrest in the Western world. In the United States, civil rights groups protested against systemic racism and discrimination, women campaigned for reproductive rights and there were widespread protests against American involvement in the Vietnam war. The patients' rights movement in the US emerged against this backdrop of social unrest [7, p. 286]. The first articulation of the standards now known as 'patients' rights' was the result of the campaigning activities of the National Welfare Rights Organisation, a grassroots organisation of poor (mostly black) women and their children whose communities had been ravaged by economic deprivation and racial discrimination. During the 1960s and early 1970s, the NWRO agitated for fairer access to healthcare and patients' rights in a hostile environment in which patienthood was a privilege [7, p. 286]. Responding in 1969 to calls from the Joint Commission on the Accreditation of Hospitals (JCAH) for feedback on newly developed quality standards for healthcare, the NWRO firmly inscribed the patient perspective on the draft standards. At the same time, the NWRO transformed the American Hospital Association's initiative to develop a mission statement for member hospitals into what became the first AHA Patient Bill of Rights in 1973 [7, p. 288]. The document represented "a near-revolutionary departure from traditional Hippocratic beneficence" – a shift from the familiar emphasis in medical codes of conduct on physicians' obligations or virtues to the language of rights and legitimate entitlement [8, p. 94]. In compelling physicians to recognise the rights of patients in the sphere of medical decision-making, the Bill of Rights of 1973 and similar documents "literally invite[d] the replacement of the beneficence model with the autonomy model" [8, p. 95]. Recognition of the human rights of the medically indigent and greater involvement of patients in the healthcare system were just two components of the NWRO's contribution to the welfare rights movement, but, in tandem with developments in the emerging field of research ethics, they drew attention to the moral obligations of doctors generated by the moral rights of patients [8, p. 95]<sup>4</sup>.

<sup>3</sup> In emphasising the importance of informed consent, the Declaration of Helsinki, which first appeared in 1964, recognised the rights of individuals to deliberate and decide without coercion or manipulation whether or not to participate in research. However, none of the early versions of the document mentioned autonomy as such. Only in 2008 did 'self-determination' first appear as one of the 'basic principles' governing research. See World Medical Association *Declaration of Helsinki* (2008). Available at <http://ethics.iit.edu/ecodes/node/4618>.

<sup>4</sup> The gradual recognition of the importance of autonomy was the product of related developments in law, medical education and research. Faden and Beauchamp provide a comprehensive account of the emergence of autonomy as a principle of medical ethics in

### Research misconduct

Several commentators link the initial appearance of the concept of autonomy in bioethics to public and professional concerns about violations of research ethics and integrity. Albert Jonson sees the emphasis on autonomy in contemporary bioethics as "stem[ming] in large part from debates over the ethics of human experimentation and from a concern about the 'excessive paternalism' exercised by physicians" [9, p. 10]. At no point in the second half of the twentieth century was this concern more pronounced than in the aftermath of the Tuskegee study of untreated syphilis in the negro male, closed down in 1972 by the US Department of Health, Education and Welfare after information leaked by whistle-blower Peter Buxton made front-page headlines in the *New York Times*. Between 1932 and 1972, researchers in Macon County, Alabama, had observed the effects of untreated syphilis on 399 impoverished African-American sharecroppers, without ever informing the participants that they were involved in a research study<sup>5</sup>. Not only did the investigators fail to obtain informed consent, they also actively took steps to prevent subjects from seeking treatment (penicillin had become available as a treatment for syphilis in the late 1940s) [10, p. 21]. By the time the study was halted, at least 28, but possibly as many as 100 subjects had died from syphilis-related complications. Funded by the US Public Health Service and approved for continuation by the Centres for Disease Control as late as 1969 [10, p. 21], the use of 'human guinea pigs' in the Tuskegee study generated outrage among members of a public newly sensitised to questions of rights, equity and discrimination [11]. To address the erosion of public trust engendered by these myriad ethical failures, President Nixon signed the National Research Act into law in 1974, creating the National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research. In the absence of legislation governing the conduct of human participant research, the remit of the Commission was to protect participants and ensure oversight of federally funded research<sup>6</sup>.

### The principlist account of autonomy

Autonomy first came to prominence in the context of biomedicine in the Belmont Report, published in 1978 by the

their historical and conceptual analysis of the doctrine of informed consent, *A history and theory of informed consent*.

<sup>5</sup> In total, 399 men infected with the disease and 201 uninfected controls were enrolled in the study without obtaining consent. Participants were informed only that they would receive free 'treatment' for 'bad blood' – a term for syphilis used colloquially in the American south (See Brandt, 1978:24).

<sup>6</sup> Tuskegee was not the only scandal which vitiated research in the US in the mid-twentieth century; other historic abuses of participants occurred at the Willowbrook State School in the late 1950s and in the Jewish Chronic Disease Hospital in the early 1960s. Although oversight of research had been an issue for executives at the National Institutes of Health from the late 1950s onwards, risk and liability were the focus of attention; neither informed consent nor protection of autonomy were prominent themes in any of the reports published during this period. See Faden, R and Beauchamp T. *A history and theory of informed consent*. Oxford: Oxford University Press (1986), 207.

National Commission for the Protection of Human Subjects of Biomedical and Behavioural Research. In the wake of the Tuskegee scandal, one of the responsibilities of the Commission was to identify the basic ethical principles which should govern the conduct of research involving human participants ([12], Section 202 (a)). Between 1974 and 1978, the Commission published 17 reports and appendix documents relating primarily to the conduct of research involving what would now be termed 'vulnerable' populations, including fetuses, prisoners, children and adults lacking decision-making capacity. According to Faden and Beauchamp, in compiling these reports, the National Commission "pursued issues of autonomy, informed consent and third-party consent more vigorously than had any previous body" [8, p. 215]<sup>7</sup>. Prior to the impetus provided by the National Commission, they argue, the ethics community had been 'napping' [8, p. 93]. The Belmont Report proposed three ethical principles for the conduct of research:

- the principle of respect for persons;
- the principle of beneficence;
- the principle of justice.

In the report, the principle of respect for persons involves the twofold claim that individuals should be treated as 'autonomous agents' and that those whose autonomy is diminished should be afforded protection. The report defines an autonomous person as "an individual capable of deliberation about personal goals and of acting under the direction of such deliberation" [13]. This definition is consonant with the account of autonomy put forward in the first edition of the seminal *Principles of biomedical ethics*, published in 1979. It is no coincidence that Tom Beauchamp, one of the authors of the *Principles*, served as a consultant to the Commission during the period in which the Belmont Report was produced. Autonomy is defined in the first edition of the *Principles* as

*"a form of personal liberty of action where the individual determines his own course of action in accordance with a plan chosen by himself or herself. The autonomous person is one who not only deliberates about and chooses such plans but who can act based on such deliberations"* [14, p. 56].

Characteristic of this definition is an emphasis on individual traits such as independence and self-reliance, which have both psychological and social dimensions. Early editions of the *Principles* propose a 'general idea of autonomy' understood in terms of "being one's own person, without constraints [imposed] by another's action or by a psychological or physical limitation" [14, p. 56–7].

The concept of self-legislation at work in the *Principles* is explicitly Kantian. Autonomous persons act on the basis of reasons which are 'their own', and these are "principled

<sup>7</sup> Faden and Beauchamp contend that, "more decisively than any previous publication in case law or research ethics, the Commission's volumes reflected the view that the underlying principle and justification of informed consent requirements, at least for autonomous persons, is a moral principle of respect for autonomy, and no other". See Faden, R and Beauchamp T. *A history and theory of informed consent*. Oxford: Oxford University Press (1986), 216.

rather than arbitrary reasons" [14, p. 58–9]. This notion of "self-directed action based on a rational principle accepted by the agent" [14, p. 56–7] is central to the principlist account of autonomy. Beauchamp and Childress attribute to Kant the idea that everyone "must make (author or originate) his own moral principles" and redefine this to mean that "each individual must will the acceptance of his principles" [14, p. 57]. However, they provide no further account of the nature of willing or of the role of principles in human action, both of which are central to the Kantian conception of moral agency.

From the third edition of the *Principles* onwards, the principle of autonomy is recast as the principle of respect for autonomy, construed as a 'moral limit,' which "obligates agents to respect the autonomy of others" [15, p. 13–4]. Justification for the 'moral prescription' to respect the autonomy of others rests squarely on what Beauchamp and Childress refer to in the second edition as the Kantian principle of respect for persons: the idea that we should respect others "as persons with the same right to their choices and actions as we have to our own", because everyone is "rightfully a determiner of his or her destiny" [16, p. 63]. On this reading, the mandate to respect autonomy "flows from the recognition that all persons have unconditional worth, solely as ends in themselves determining their destinies" [14, p. 57–8; 16, p. 62]. Failing to respect autonomy, conversely, involves treating an individual "merely as a means, because he or she is treated in accordance with rules not of her choosing" [16, p. 62]. The Kantian distinction between 'ends in themselves' and 'mere means' which underpins the principlist defence of autonomy is not further explored in the text. I will return to this issue below.

### Positive and negative liberty

The principlist definition of autonomy is indebted not only to Kant, but also to the utilitarian philosopher John Stuart Mill, a debt explicitly acknowledged by Beauchamp and Childress [14, p. 56–7]<sup>8</sup>. Although conventional histories of moral philosophy represent the respective moral theories of Kant and Mill as irreconcilable<sup>9</sup>, it is not Mill's ethics, but his political theory, which is of interest to Beauchamp and Childress. Anchoring human well being in the 'free development of individuality' [17, p. 57], Mill argues that "there is a limit to the legitimate interference of collective opinion with individual independence" [17, p. 8–9]<sup>10</sup>. While the

<sup>8</sup> In the second edition, the principle of autonomy is derived both from Mill's view that "insofar as an autonomous agent's actions do not infringe upon the autonomous actions of others, that agent should be free to perform whatever action he or she wishes" and from the Kantian doctrine of respect for persons. See Beauchamp T. and Childress J. *Principles of biomedical ethics* (second edition). Oxford: Oxford University Press (1983):62–3.

<sup>9</sup> For Kant, the moral worth of an action lies in the reason or intention underlying the action, whereas for Mill, actions are judged morally praiseworthy or blameworthy by reference to their consequences (whether or not they lead to an increase in overall happiness – the so called 'utility principle'). There is no point of reconciliation between these two viewpoints.

<sup>10</sup> Although Mill rarely used the term autonomy, O'Neill argues that his writings on liberty can be seen as "the most profound attempt

fundamental principle of political liberalism is the protection of a “minimum area of personal freedom which must on no account be violated” [18, p. 124]<sup>11</sup>, Mill also insists on the importance of what he terms ‘character’: the ability to “[make one’s] desires and impulses [one’s] own” or to allow one’s desires to express one’s own nature [17, p. 60].

In 1958, Isaiah Berlin developed this insight by distinguishing between the ‘negative’ freedom to “act unobstructed by others” [18, p. 122] and a ‘positive’ conception of liberty defined in terms of a more internal form of independence understood as self-mastery [18, p. 131]<sup>12</sup>. Contemporary discussions of autonomy and of agency owe much to Berlin’s conception of positive liberty, described in the following terms:

*“I wish to be a subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me, as it were, from the outside. I wish to be somebody, not nobody: a doer—deciding, not being decided for, self-directed and not acted upon by external nature or by other men. (...) I wish, above all, to be conscious of myself as a thinking, willing, active, being, bearing responsibility for my choices and able to explain them by reference to my own ideas and purposes”* [16, p. 131].

The principle of respect for autonomy put forward by Beauchamp and Childress incorporates elements of both positive and negative conceptions of liberty. In the second edition of the *Principles*, the principle of autonomy is stated as the principle that “[a]utonomous actions and choices should not be constrained by others” [16, p. 62], invoking Feinberg’s notion of a “realm of inviolable sanctuary [which] most of us sense in our own beings” [19, p. 27]. Yet, as implied above, freedom from external interference is only a necessary, but not a sufficient condition of autonomy, the exercise of which presupposes an additional range of capacities, including the “ability to decide for oneself, control one’s life and absorb the costs and benefits of one’s choices” [20]. As Gaus points out, the notion of a free person as someone whose actions are in some sense her own is a central theme of liberal political philosophy [21]. The ideal rational chooser of the *Principles* is presumably Berlin’s positively free subject, moved by reasons which are her own, and assuming responsibility for her choices. To be one’s own person in this sense, however, is “to be directed by considerations, desires, conditions, and characteristics that are not simply imposed externally upon one, but are part of what can somehow be considered one’s authentic self” [20]. The exercise of autonomy consists, not merely in exercising choice, but in “the condition of being self-directed, of hav-

ing authority over one’s choices and actions whenever these are significant to the direction of one’s life” [22, p. 100]. It is a significant flaw in the principlist account of autonomy that it fails to address these vital questions of authority and authenticity<sup>13</sup>.

## Back to Kant

Although current discussions of autonomy are prefigured by JS Mill’s notion of character [17, p. 60] and John Rawls’ conception of autonomy as the product of moral education in a well-ordered society [23, p. 514] – it is without question Immanuel Kant whose conception of autonomy has exerted the most profound influence on contemporary debates. Kant, unlike contemporary theorists, saw autonomy as ineliminably bound up with morality. The conception of morality as autonomy invented by Kant [24, p. 3] was built on the enlightenment understanding of morality as self-governance – an ideal put forward in opposition to earlier interpretations of morality as obedience [24, p. 4]. Carving out a conceptual space for a social sphere “in which we may each rightly claim to direct our actions without interference from the state, the church, the neighbours or those claiming to be better or wiser than we”, this new conception of morality signalled the emergence of a distinctive Western liberal view of the “proper relations between individual and society” [24, p. 4–5].

## The rationality of morality

Kant’s philosophical system was built on a strict distinction between the world of empirical causes and effects and the realm of morally motivated action. Kant was both a scientist and an ethical theorist: he sought to explain in terms congruent with Newtonian physics how we can know the natural world, but he also wanted to explain moral motivation in a way which didn’t reduce morality to a form of empirical causality<sup>14</sup>. To do this, he needed a framework distinct from his account of how we understand the behaviour of the natural world. Morality, for Kant, is a special kind of causality: a force capable of bringing about effects in the natural world which originates outside of the chain of empirical causes and effects<sup>15</sup>. It originates, in other words, in the human capacity to will – the capacity to “make things happen intentionally and for reasons” [25, p. 18], or, in other words, to recognise incentives or motives or

to set autonomy within a naturalistic account of human action”. See O’Neill, Onora. *Autonomy and trust in bioethics*. Cambridge: Cambridge University Press (2002), 30.

<sup>11</sup> For Mill, the only justification for state interference with this domain of freedom is the prevention of harm to others (Mill, 13).

<sup>12</sup> Strictly speaking, it was Kant who first made the distinction between positive and negative freedom in 1785. See Kant, Immanuel. *Groundwork of the metaphysics of morals*, Cambridge: Cambridge University Press (1997), 53–4.

<sup>13</sup> Interestingly, Faden and Beauchamp discuss the issue of authenticity in some detail in *A history and theory of informed consent*, but it is not dealt with in the *Principles*. See Faden, R and Beauchamp T (1986). *A history and theory of informed consent*. Oxford: Oxford University Press, 235–273.

<sup>14</sup> Kant was an enlightenment thinker and a perfectionist; his aim in developing his moral philosophy was not to convert those who were sceptical about the possibility of an ‘objective’ foundation for morality, but to persuade eighteenth-century readers who viewed themselves as rational that morality had a rational basis.

<sup>15</sup> In contemporary parlance, Kant was a compatibilist: he maintained both that, as part of nature, humans are determined by the causal laws which govern the behaviour of all natural things, and also that, as rational beings, we are free to will and act independently of the laws of nature. In short, we are both determined and free.

principles as sufficient to move us to action. Without this distinctive capacity, there can be no moral responsibility: human action would be determined by the same laws which dictate plant and animal behaviours and geophysical events. While it would be inaccurate to describe Kant as a rationalist, his construal of human beings as essentially rational creatures is fundamental to any understanding of his moral theory<sup>16</sup>.

Autonomy, for Kant, is a defining characteristic of the human will – or rather, of the will of human beings who are capable of rational deliberation<sup>17</sup>. All actions involve willing; to will something is to select a general principle which expresses what one intends to achieve by performing a given action<sup>18</sup>. Kant sees the predictable relationship between intention and outcome in terms of lawlikeness; in every case, if our reason for acting is sufficient to motivate us, we take action. Human willing, in other words, has a structure: it cannot be totally random. Otherwise, we would not be able to evaluate our own or anyone else's reasons for acting [26, p.15] – an evaluation which is presupposed by morality. Central to Kant's conception of morality is the basic idea that moral action involves obeying a law which our reason gives to itself<sup>19</sup>. The law must come from reason because the authority of reason "hold[s] indifferently for all rational beings" [27, p. 54]. Self-governance, for Kant, is governance by reason.

Each individual human being has her own desires, appetites, wants and interests, formed over the course of a lifetime; these are what distinguish us from other human beings. While most of the actions we perform are based on these individual needs and desires, from a rational perspective, these are 'arbitrary' because their validity or acceptability is limited to a single individual. Moral action, by contrast, requires abstracting from this biased point of view to attain a more universal perspective. Self-legislation, for Kant, rests on the idea that moral action is based on principles which are not arbitrary or based on a particular individual's contingently-formed needs or appetites; these principles have the higher authority of reason and as such they are potentially acceptable to all other agents<sup>20</sup>. In other words, when it comes to moral action, an agent's point

of reference in willing a given action is not her particular wants or interests as an individual, but the rational interests of human beings as such. Kant's 'categorical imperative' or moral law – "Act only on that maxim which you can at the same time will to be a universal law" – is the product of his search for a rational principle which was "simple enough to be known by and accessible to everyone and clear in its motivational force" [24, p. 10–11].

The concept of autonomously willed action as action based on principles which are valid for all agents, not just for some agents or for a single individual [26, p. 16] has its roots in the idea of the general will put forward by Rousseau in the *Social contract*. According to Rousseau – who was no rationalist – by agreeing to enter into a social contract with others, man trades the freedom he had in the state of nature for the 'moral freedom' which he acquires as a result of becoming part of civil society [28, p. 53–4]. 'Natural' freedom here is the freedom to act on one's impulses and appetites without constraint, whereas moral freedom "alone makes man truly the master of himself" [28, p. 54]. This is because the general will – the will of the body politic – "always tends towards the preservation and welfare of the whole and of each part" [24, p. 474]. In accepting the general will as their own will, members of civil society see themselves a part of a 'moral whole' and commit themselves to the pursuit of a common good, but do so in the knowledge that they "must limit [their pursuit of their own interests] to those compatible with everyone else's pursuit of their acceptable ends" [24, p. 495]. Autonomy, on this view, is "legislation by coordinated individual selves" [26, p. 12] and, unlike natural freedom, which is 'accidental', the freedom which is conferred by this legislation, is an essential component of being human [29, p. 478]. Ultimately, for Rousseau, as for Kant, "the impulsion of mere appetite is slavery" [28, p. 54]; unlike Rousseau, however, Kant maintained that freedom could be attained by the individual independently of society.

Kant understands autonomy as "that property of the will by which it [the will] is a law to itself" [30, p. 47] and moral action as "the subjection of reason to no laws except those it gives itself" [31, p. 15]. Human beings are free in the sense that they can choose whether or not to exercise their autonomy, but they are ultimately unfree if they choose not to exercise their autonomy. This paradox is rooted in Kant's dualistic conception of human nature. While human nature, for Kant, is both rational and animal, there is a clear hierarchy; desires, appetites, impulses and emotions originate in the animal (or 'sensible') part of our nature and as such they are 'alien' to our rational nature [30, p. 62]. Actions based on these elements of human nature are a product of what Kant terms 'heteronomy' and as such can never be autonomous [30, p. 47]<sup>21</sup>; we acknowledge their alienness by ruling them out as legitimate reasons for acting.

in principle be possible for all other agents acting in similar contexts to accept their validity.

<sup>21</sup> "If the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law (...) heteronomy always results". See Kant, Immanuel. *Groundwork of the metaphysics of morals*, Cambridge: Cambridge University Press (1997), 47.

<sup>16</sup> What distances the contemporary reader most from Kant's moral theory is his 'apriorism', a technical term for his insistence that moral motivation is something which we can never acquire from experience because it originates in our rational nature, independently of, and prior to, any actual experience.

<sup>17</sup> Although Kant does not broach the subject of human beings who are incapable of rational deliberation, it is not likely that he would have attributed autonomy to them.

<sup>18</sup> For Kant, agents have an understanding of what they are doing and why; this can be expressed in what he calls a 'maxim' of action. See Hill, Thomas. "Kantian autonomy and contemporary ideas of autonomy", in (ed) Sensen. *Kant on moral autonomy*. Cambridge: Cambridge University Press (2013), 18.

<sup>19</sup> To underscore the universality and necessity of the moral principles he puts forward, Kant conceives them as laws; obligations based on these principles are as exceptionless in the realm of human action as their analogue, the laws of nature, are in the empirical world.

<sup>20</sup> This is what Kant means when he says that the maxims (or general principles) on which we act must be 'universalisable'; it would

Unlike autonomously willed actions, actions based on 'heteronomous' reasons "are ultimately defective, incomplete or 'private' reasons" [26, p. 17]. Simply put, moral action cannot be based on a reason which would not be acceptable to any other rational agent [26, p. 12]. The fundamental idea here is that we express our rationality – and express respect for our own and others' rational nature – by exercising autonomy; this necessarily involves an understanding of ourselves as moved to act solely by motives which are rational<sup>22</sup>. As a motivating force, reason is more compelling than instinct for beings like us who understand themselves as both animal and rational<sup>23</sup>. And reason can trump instinct precisely because the human will is autonomous<sup>24</sup>.

Closely linked to the idea of respect for our rational nature is the idea that human beings must always be regarded as ends in themselves, not merely "means to be used by this or that will at its discretion" [30, p. 37]. Unlike the objects of our contingent desires which have only conditional value (conditional upon their satisfying a given desire), human beings qua rational beings have absolute or unconditional worth. "Rational beings are called persons, because their nature marks them out as an end in itself, that is, as something that may not be used merely as a means (...) (and is an object of respect)" [30, p. 37]. What is wrong with treating someone merely as a means is not simply that that person's own ends or goals are ignored; more precisely, it involves treating her "in ways that are appropriate to things", and this negates her autonomy by damaging her capacity for rational agency [27, p. 138]. This unequivocal prohibition on the instrumentalisation of human beings is what Kant means by 'respect for persons' but it must be understood in the broader context of Kant's account of rational willing; what is valuable for Kant and worthy of respect is precisely the activity which constitutes the exercise of rational agency: reflection and discrimination in relation to one's motives and reasons for action.

In sum, autonomously willed actions are the product of a process of deliberation which involves scrutiny of the rea-

<sup>22</sup> Agents can ensure that their reasons for action are uncontaminated by 'alien' elements by performing a thought experiment which tests the rationality of their reasons for acting. The so-called categorical imperative – "Act only on that maxim which you can at the same time will to become a universal law" – states that, if the reason or principle on which an action is based could be accepted by any other rational being in identical circumstances, there is no inconsistency in the thought that it could function as a universal law or principle, obligating all rational agents to do the same thing.

<sup>23</sup> Kant states that "nothing other than the representation of the [moral] law itself, which can of course only occur in a rational being (...) can constitute the preeminent good we call moral, which is already present in the person himself who acts in accordance with this representation and need not wait upon the effect of his action" See Kant, Immanuel. *Groundwork of the metaphysics of morals*. Cambridge: Cambridge University Press (1997), 14.

<sup>24</sup> An autonomous will is a necessary but not a sufficient condition of moral action for Kant. In addition to being uncontaminated by motives 'alien' to the rational self, an action must be motivated by a desire to do what is right – acceptable to other rational beings – because it is the right thing to do, and because it expresses our rational nature. Kant, Immanuel. *Groundwork of the metaphysics of morals*. Cambridge: Cambridge University Press (1997), 14.

sons underlying our proposed actions to ensure that they meet certain standards of rationality. Thomas Hill contemporarises Kant by attributing to him less a metaphysical view about human nature than "a normative idea about the task, attitudes and commitments of rational agents when deliberating about what to do" [32, p. 84]. Although Kant is primarily interested in the nature of moral motivation, he is also concerned with responsibility and with our understanding of what counts for us as sufficient reasons for acting [32, p. 86]. If it is indeed the case that "the common core of all sorts of heteronomous willing is that it is not fully reasoned", we must concede that in these instances we exercise insufficient authority over our actions [25, p. 10]. In rational deliberation, conversely, we take ourselves to be able to resist being causally determined by desire or impulse, but we also take ourselves "to be able to act for the sake of ends other than the satisfaction of desire" [32, p. 86]. Descended from this Kantian account is the more contemporary understanding of an autonomous agent as someone who has the ability to "deliberate about action and to author action (...) [but who is also] free to change the values and motivations about which she deliberates and to alter her life activities if she so chooses" [33, p. 185]. Autonomy is fundamental to this understanding of ourselves as the kinds of beings who can canvass our reasons for acting, recognising some as more compelling than others.

## Autonomy: the next generation

Like 'freedom', 'autonomy' is a term "so porous that there is little interpretation that it seems able to resist" [17, p. 121]. The remainder of this paper will focus on contemporary discussions of personal autonomy, with the caveat that even this narrower concept is fraught with ambiguity. In *The theory and practice of autonomy*, Gerald Dworkin identifies nine synonyms for personal autonomy, five sets of capacities suggested by the exercise of autonomy and seven types of object related to its employment [34, p. 6]. Perhaps more helpfully, Sarah Buss divides contemporary theories of personal autonomy into four 'overlapping' families, all of which presuppose that agents can become aware of the beliefs and desires that move them to act, since "self-government requires two points of view: that of the governing authority, and that of the governed" [35]. The first family of theories emphasises the coherence between the reason an agent selects as the basis for her action and her overall attitude towards that action; to be autonomous, on this view, an action must not be discordant with the agent's sense of who she is [35]. A second family of theories construes the autonomous agent as responsive to a variety of reasons for and against performing the proposed action; the agent 'authorises' her action based on an understanding that the reason she selects as the basis for her action is more compelling than any of the available alternatives [35]. A third family of theories stresses the agent's ability to evaluate her reason for performing a given action in the light of her other beliefs and desires, adjusting her actions if they are not supported by her existing beliefs and values. A fourth family emphasises freedom from determination by



external causes<sup>25</sup>. While the minutiae of the distinction between these different accounts of autonomous action is perhaps of interest only to specialists, this categorisation is useful in the sense that it makes explicit several important aspects of autonomy absent from the principlist account. I will examine two of these below.

### Reflective endorsement

Theories which rely on the idea of coherence between one's reason for acting and one's sense of self involve the claim that an agent who is free or autonomous identifies with her reason for performing a certain action; if she is alienated from her reason for acting she cannot be said to be free. This is the position represented by the 'hierarchical' theories of Harry Frankfurt and Gerald Dworkin. For Frankfurt, what is distinctive about humans is their facility for reflectively evaluating the desires and motives that they have. Humans can form a hierarchy of desires, incorporating not only straightforward 'first order' desires for one thing or another, but also 'second-order' desires, which are essentially desires to "want to have (or not to have) certain [first-order] desires" [36, p. 6–7]. We can have desires, in other words, which we override because we don't want them to be 'effective' in motivating us to act. For Frankfurt, it is only because we have this capacity to form second-order desires that we can be said to have free will [36, p. 19]. Someone who is free in this sense is free, not to do as she wants, but to want what she wants to want [36, p. 21]; she exercises freedom of the will (or autonomy) when her will 'conforms' to her second-order desires [36, p. 21]. For Frankfurt, "it matters greatly to us whether the desires by which we are moved to act as we do motivate us because we want them to be effective in moving us or whether they move us regardless of ourselves or even despite ourselves" [37, p. 163]. If we don't 'unequivocally endorse' the motive from which we act, it can't be said to be something we really want. When we act on the basis of a desire which we don't really want to have, we are passive because we are moved to act "by a force that is not fully our own" [37, p. 164].

Like Frankfurt, Gerald Dworkin maintains that it is the capacity to form attitudes towards the reasons which motivate us to act which is definitive of human freedom<sup>26</sup>. We can either identify with these influences, wishing to be motivated by them, or we can be alienated from them, wishing to be motivated in different ways [38, p. 24–5]. We are not unfree if we do something we don't want to do, as long as we identify with the reason for our action [39, p. 377] – an agent may perform an action she has no desire to perform, for example, because she wants to be able to see herself as a certain kind of person. The capacity for autonomy is not merely a capacity to reflect on and evaluate one's desires or preferences, however, but also an ability "to alter one's preferences and to make them effective in

one's actions, and indeed, to make them effective because one has reflected upon them and adopted them as one's own" [34, p. 17]. Neither Frankfurt's nor Dworkin's theory of autonomy imposes any restrictions on the content of what an agent can desire; both are 'procedural' accounts of autonomy which emphasise the conditions in virtue of which an action can be considered autonomous [40, p. 125]. Although there are problems with hierarchical theories<sup>27</sup>, their value lies in the fact that they draw attention to "a very special ability – [the ability] to rise above or step back from our own attitudes, reflectively evaluating them and forming higher-level attitudes to them" [41, p. 147]. This activity of reflective endorsement "ensures [that] our desires don't automatically motivate us to act" [41, p. 160].

### Authenticity

Because hierarchical theories of the kind described above fail to explain why a higher-order desire is more authoritative than a lower-order desire, another family of theories emphasises an agent's understanding of how she came to have a particular desire or preference. For John Christman, a central component of autonomy is the extent to which the agent participates in the process of how her preferences are formed and whether she could have 'resisted' the formation of the desire she proposes to act upon, if she had been given the opportunity [42, p. 10]. On this view, an autonomous agent is capable of becoming aware of the ways in which her character changes and develops and of understanding why these changes come about; this self-awareness allows her to facilitate or to resist such changes. Presuming that the agent is not labouring under some form of self-deception, the capacity for reflection and the capacity for self-awareness are preconditions of a transparent process of preference formation [42, p. 10]. A person chooses or judges 'for herself' only when she "is in tune with the settled aspects of herself that apply to the judgment at hand" – only when the desires and beliefs motivating her at a particular point in time are 'transparent' to her [42, p. 17]<sup>28</sup>. Conversely, "[i]f the 'self' doing the 'governing' is dissociated, fragmented, or insufficiently transparent to itself, then the process of self-determination sought for in a concept of autonomy is absent or incomplete" [42, p. 17].

Rather than endorsement of actions which affirm a person's sense of self-identity, Marina Oshana claims that what is required for autonomy is that an agent be true to herself in the sense of being "disposed to acknowledge the aspects of her character and history that anchor or contradict her self-conception" [3, p. 19]. This emphasises the integration of aspects of one's self or character over time, rather than integration of higher and lower-order desires at a given point in time. For Oshana, autonomy requires "having a sense of

<sup>25</sup> In philosophical parlance, these are known, respectively, as 'coherentist' theories, 'reasons-responsiveness' theories, 'responsiveness-to-reasons' theories and 'incompatibilism'.

<sup>26</sup> Neither Frankfurt nor Dworkin use the term 'autonomy' in their writings from the early 1970s; what they mean by 'freedom' in these early essays is identical to what is referred to by later commentators as 'autonomy'.

<sup>27</sup> The main criticism levelled against hierarchical theories is that there is no evident reason why second-order desires should not in turn require endorsement at a higher level, and so on, leading to an infinite regress.

<sup>28</sup> This presupposes that the autonomous person is "not being guided by manifestly inconsistent desires or beliefs". See Christman, John (1991) "Autonomy and personal history". *Canadian Journal of Philosophy* (March 1991) 21 (1), 17.

what is important to oneself” [43, p. 243] and in practical terms being able to make choices which express who one is. The autonomous agent “must have the power to deliberate about and to change her values and motivations and to alter significant relations in her life if she so chooses” [33, p. 198]. This requires, not just the freedom to decide or to act, but also the ability to own or authorise our decisions and actions – having what Oshana terms “agential authority over these decisions and actions” [33, p. 199]. At issue here is not endorsement of desires or actions so much as an understanding of what we accept as authoritative in motivating ourselves to act [33, p. 198].

### Philosophy and bioethics

The question which remains to be answered is whether any of these more exacting accounts of autonomy have any greater value for bioethical analysis than does the principlist definition. In more recent editions of the *Principles*, Beauchamp and Childress explicitly repudiate the claim that autonomy requires the capacity to identify with or reflectively endorse one’s desires or preferences, arguing that higher-order desires are not automatically more authoritative than lower-order desires [44 p. 100–1]. From their perspective, the problem with hierarchical or ‘coherentist’ theories is that the requirement to reflect on one’s preferences at a higher level constitutes an ‘aspirational ideal of autonomy’ which is “beyond the reach of normal agents and choosers” [44, p. 101]. Beauchamp and Childress defend a weaker conception of autonomy – or rather, autonomous action – which applies in ‘nonideal’ conditions to the actions of ‘normal choosers’ who “act intentionally, with understanding and without controlling influences” [44, p. 101]. They claim to provide a theory of autonomy which is “consistent with pretheoretical assumptions implicit in the principle of respect for autonomy”, rather than a ‘mythical ideal’ which disqualifies the routine choices of patients and research participants from counting as autonomous. According to this ‘pretheoretical’ model, to respect autonomous agents involves acknowledging their right “to hold views, to make choices and to take actions based on their personal beliefs and values” [44, p. 103]. Yet this assumption requires some theoretical unpacking: to make choices – at least important choices – involves applying a principle of selection and an understanding of why one option was preferred to another, and this understanding is to some degree informed by what Ronald Dworkin refers to as one’s ‘critical interests’ [45, p. 201]. And to base one’s actions on one’s beliefs and values requires at least the ability to make explicit the connection between one’s actions and one’s beliefs and values, which in turn presupposes sufficient self-awareness to identify the values which one considers important (and possibly why).

Beauchamp and Childress are to be commended for placing autonomy as an ideal of non-instrumentalisation at the heart of bioethics discourse after Tuskegee. In essence, the principle of respect for autonomy states that it is immoral to exploit, manipulate or coerce others because humans capable of deliberation and self-determination should not be the means to others’ ends. But a more philosophically robust account of autonomy is needed to justify this principle. Ultimately, it is not clear what distinguishes their account

from “an affirmation of choice itself, where all options are equally worthy, because they are freely chosen, and it is choice that confers worth” [46, p. 36–7]. Moreover, their conception is too easily reduced to the procedural components of informed consent. According to O’Neill, “[w]hat is rather grandly called patient autonomy often amounts simply to a right to choose or refuse treatments on offer, and the corresponding obligations of practitioners not to proceed without patient’s consent” [1, p. 37]<sup>29</sup>. In fact, the translation of the concept of autonomy into consent law “highlights the dissonance between the ethical importance of respect for autonomy and the fundamental goals and processes of the law” [47, p. 86]. Since providing consent need not be “accompanied by a ‘high degree of reflection’,” autonomous choice is not guaranteed by the informed consent process, and ultimately it may not even be required by this process [1, p. 37].

### Conclusion

What is missing from the principlist account is the reflexivity central to historical and contemporary philosophical theories of autonomy. I would argue that this ultimately undermines the value of autonomy where it matters: in the clinical contexts in which patients’ value-systems and what Oshana calls ‘central and unsheddable’ elements of their identity play a critical role in their decision-making [3, p. 7]. In these specific situations, a reflective understanding of one’s preferences and values and the roles these play in contributing to one’s identity is required if autonomy is not to collapse back into what O’Neill refers to as ‘mere sheer choice’ [48].

For Beauchamp, “everyday choices of ordinary persons are paradigm cases of autonomous choices” and applying more demanding standards of reflexivity would render many of these quotidian decisions non-autonomous [49, p. 313]. This is a concern first articulated by Faden and Beauchamp in 1986: the idea that ‘normal choosers’ in most situations where decisions are required would “fail to engage in higher-order reflection” [49, p. 319] and would not therefore be ‘deserving of respect’ for their autonomous choices [44, p. 101]<sup>30</sup>. Faden and Beauchamp consider but explicitly repudiate the idea of authenticity – understood in terms of consistency between a person’s action and her ‘reflectively accepted values’ – as a condition of autonomy, because it is ‘overly demanding’ [8, p. 264]. There are two points to be made here. First, perhaps Faden and Beauchamp exaggerate the overdemandingness of the authenticity requirement and expect too little of ‘normal choosers’. Most people, faced with important decisions, are capable of reflection on the relationship between their values, their sense of self and the options available to them. As Dworkin points out, even in the

<sup>29</sup> “What passes for patient autonomy in medical practice is operationalised by practices of informed consent: the much-discussed triumph of autonomy is mostly a triumph of informed consent requirements”. See O’Neill, Onora. *Autonomy and trust in bioethics*. Cambridge: Cambridge University Press (2002), 38.

<sup>30</sup> See Faden, R and Beauchamp T. *A history and theory of informed consent*. Oxford: Oxford University Press; 1986. 235–7.

absence of extensive education, it is possible for someone to 'shape and mould' her life – to form opinions, to change – in accordance with "reflective procedures" [17,34]. Secondly, however, very few of our daily decisions require this level of reflection or present us with an opportunity to interrogate our preferences and evaluate the values on which we propose to base our actions. For the most part, we don't have cause to reflect on what motivates us in our everyday choices, unless we have a deeper interest in the outcome of those decisions. But some decisions do require this attention to the coherence between preferences, values and something like a 'settled' sense of self, and many of these arise in the healthcare setting. Consider, for example, an independent elderly bachelor choosing between discontinuing dialysis and a lower-leg amputation, or a couple faced with a prenatal diagnosis of fetal abnormality, or a woman thinking about undergoing IVF. These decisions are the ones which demand respect for autonomy in the strong sense; what is called for in these instances is not simply non-interference and the provision of information, but active measures to enhance a person's capacity to understand the reasons motivating her to act and to appreciate the coherence of those reasons within her overall view of herself and her place in the world.

What I am suggesting here is that, in bioethics discourse, we should use the term autonomy with caution, applying it to situations in which it can meaningfully illuminate decisions or guide practice. It should not be applied to the routine decisions which are the target of Beauchamp and Childress' analysis; instead, the focus of these kinds of decisions should be on enhancing the process of informed consent. It should be applied to decisions which engage someone's "most deep-seated commitments, desires and preferences" [49, p. 320], based on "values, desires and convictions that have developed in an uncoerced and conscious fashion" [50, p. 82]. Autonomy understood merely as the uncoerced, informed exercise of choice undermines its powerful symbolism: the idea that this unique capacity for reflection defines us as human beings, imposes a universal prohibition on instrumentalisation and enjoins on us a duty to advocate for the interests of those in whom this capacity is lacking.

## Disclosure of interest

The author declares that she has no competing interest.

## References

- [1] O'Neill O. *Autonomy and trust in bioethics*. Cambridge: Cambridge University Press; 2002.
- [2] Mason JK, Laurie GT. *Mason and McCall Smith's law and medical ethics, ninth edition* Oxford: Oxford University Press; 2013.
- [3] Oshana M. Autonomy and the question of authenticity. *Soc Theor Pract* 2007;33:411–9.
- [4] Dworkin G. Review of James Stacy Taylor, practical autonomy and bioethics. In: *Notre Dame Philosophical Reviews*; 2009. Available at: <https://ndpr.nd.edu/news/24178-practical-autonomy-and-bioethics/>.
- [5] Danner Clouser K. (1980). *Teaching bioethics: strategies, problems and resources*. Hastings on Hudson: the Hastings Center (1980), 55. Cited in (8) Faden R and Beauchamp T. *A history and theory of informed consent*. Oxford: Oxford University Press; 1986. p. 55.
- [6] Arras J. Theory and bioethics. In: Edward NZ, editor. *The Stanford encyclopedia of philosophy*. Winter 2016 Edition 2016. <https://plato.stanford.edu/archives/win/entries/theory-bioethics>.
- [7] D'Oronzio JC. A human right to healthcare access: returning to the origins of the patients' rights movement. *Camb Q Health Ethics* 2001;10:285–98.
- [8] Faden R, Beauchamp T. *A history and theory of informed consent*. Oxford: Oxford University Press; 1986.
- [9] Jonsen AR. Introduction to the history of bioethics. In: Jecker NS, Jonsen AR, Pearlman RA, editors. *Bioethics: an introduction to the history, methods and practice*. London: Jones and Bartlett Publishing; 1997. p. 3–11.
- [10] Brandt AM. Racism and research: the case of the Tuskegee syphilis study. *Hastings Cent Rep* 1978;8:21–9.
- [11] Heller J. Syphilis victims in U.S. study went untreated for 40 years. *New York Times*; 1972. Available at: <http://www.nytimes.com/07/26/archives/syphilis-victims-in-us-study-went-untreated-for-40-years-syphilis.html>.
- [12] National Research Act. *National Research Act*; 1974. Available at <https://history.nih.gov/research/downloads/PL93-348.pdf>.
- [13] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*; 1979. Available at: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>.
- [14] Beauchamp T, Childress J. *Principles of biomedical ethics, 1st ed.* Oxford: Oxford University Press; 1979.
- [15] Childress J. The place of autonomy in bioethics. *Hastings Cent Rep* 1990;20:12–7.
- [16] Beauchamp T, Childress J. *Principles of biomedical ethics, 2nd ed.* Oxford: Oxford University Press; 1983.
- [17] Mill JS. *On liberty and other writings*. Cambridge: Cambridge University Press; 1989.
- [18] Berlin I. Two concepts of liberty. In: *Four essays on liberty*. Oxford: Oxford University Press; 1969. p. 118–72.
- [19] Feinberg J. *Harm to Self (volume 3): the moral limits of the criminal law, 3*. Oxford: Oxford University Press; 1986.
- [20] Christman J. Autonomy in moral and political philosophy. In: Edward NZ, editor. *The Stanford encyclopedia of philosophy*. 2003 Ed. 2003. <https://plato.stanford.edu/archives/fall2003/entries/autonomy-moral/>.
- [21] Gaus G, Courtland SD, Schmidtz D. Liberalism. In: Edward NZ, editor. *The Stanford encyclopedia of philosophy*. Spring 2015 Ed. 2015. <https://plato.stanford.edu/archives/spr/entries/liberalism/>.
- [22] Oshana M. How much should we value autonomy? *Soc Philos Policy* 2003;20:99–126.
- [23] Rawls J. *A theory of justice*. Cambridge, Mass: Harvard University Press; 1971.
- [24] Schneewind JB. *The invention of autonomy: a history of modern moral philosophy*. Cambridge: Cambridge University Press; 1998.
- [25] Hill. Kantian autonomy and contemporary ideas of autonomy. In: Sensen O, editor. *Kant on moral autonomy*. Cambridge: Cambridge University Press; 2013. p. 15–31.
- [26] O'Neill O. Autonomy: the Emperor's new clothes. *Proceedings of the Aristotelian Society Supplementary Volume*, 77; 2003. p. 1–21.
- [27] O'Neill O. *Constructions of reason*. Cambridge: Cambridge University Press; 1989.
- [28] Rousseau J-J. In: Gourevitch V, editor. *The social contract and other later political writings*. Cambridge: Cambridge University Press; 2012.

- [29] Neuhouser F. Jean-Jacques Rousseau and the origins of autonomy. *Inquiry* 2011;54:478–93.
- [30] Kant I. *Groundwork of the metaphysics of morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press; 1997.
- [31] Kant I. What does it mean to orient oneself in thinking? (1786). In: *Religion and rational theology*. Translated by Allen Wood and George di Giovanni. Cambridge: Cambridge University Press; 1996. p. 133–46.
- [32] Hill Thomas E. *Dignity and reason in Kant's moral theory*. Ithaca and London: Cornell University Press; 1992.
- [33] Oshana M. Autonomy and free agency. In: James Stacey T, editor. *Personal autonomy: new essays of personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press; 2005. p. 183–204.
- [34] Dworkin G. *The theory and practice of autonomy*. Cambridge: Cambridge University Press; 1988.
- [35] Buss S. Personal autonomy. In: Edward NZ, editor. *The Stanford encyclopedia of philosophy*. Winter 2016 Ed. 2016. <https://plato.stanford.edu/archives/win/entries/personal-autonomy/>.
- [36] Frankfurt HG. Freedom of the will and the concept of a person. *J Philos* 1971;68:5–20.
- [37] Frankfurt HG. Identification and wholeheartedness. In: *The importance of what we care about*. Cambridge: Cambridge University Press; 1998. p. 159–76.
- [38] Dworkin G. Autonomy and behavior control. *Hastings Cent Rep* 1976;1:23–8.
- [39] Dworkin G. Acting freely. *Noûs* 1970;4:367–83.
- [40] Benson P. Feminist intuitions and the normative substance of autonomy. In: James Stacey T, editor. *Personal autonomy: new essays of personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press; 2005. p. 124–42.
- [41] Waddell Ekstrom L. Autonomy and personal integration. In: James Stacey T, editor. *Personal autonomy: new essays of personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press; 2005. p. 143–61.
- [42] Christman J. Autonomy and personal history. *Can J Philo* 1991;21:1–24.
- [43] May L. Metaphysical guilt and moral taint. In: May L, Hoffman S, editors. *Collective responsibility: five decades of debate in theoretical and applied ethics*. Lanham, Md: Rowman and Littlefield; 1991. p. 239–54.
- [44] Beauchamp T, Childress J. *Principles of biomedical ethics*, 6th ed. Oxford: Oxford University Press; 2009.
- [45] Dworkin R. Life's dominion: an argument about abortion, euthanasia and individual freedom. New York: Alfred A. Knopf; 1993.
- [46] Taylor C. *The ethics of authenticity*. Cambridge, Mass and London: Harvard University Press; 1991.
- [47] McLean S. *Autonomy, consent and the law*. Oxon and New York: Routledge; 2010.
- [48] O'Neill O. Transcript of evidence from baroness Onora O'Neill to the Commission on Assisted Dying; 2011. Available at: <http://commissiononassisteddying.co.uk/wp-content/uploads/2011/03/Baroness-O'Neill-Transcript.pdf>.
- [49] Beauchamp T. Who deserves autonomy and whose autonomy deserves respect? In: James Stacey T, editor. *Personal autonomy: new essays of personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press; 2005. p. 310–29.
- [50] Oshana M. Personal autonomy and society. *J Soc Philo* 1998;29:81–102.