

Wojciech Graboń
University of Warsaw, Poland
e-mail: wojciech.grabon@uw.edu.pl
ORCID: 0000-0002-5939-9103

DO WE HAVE AN ALTERNATIVE TO THE CRITERIALIST VIEW OF PERSONHOOD?

Abstract

The main aim of this paper is to assess various contemporary methodological approaches to the definition of personhood from the perspective of legal theory. The criterialist point of view is undoubtedly the dominant one in contemporary analytic philosophy. It assumes that our classification of entities that can be characterized as persons follows from their fulfilment of certain necessary and sufficient conditions. This kind of approach stems from a series of questions about the nature of personhood that have been analysed at least since Locke, and has found its contemporary articulation (more or less explicitly) in the work of authors such as Frankfurt, Dennett, French or Goodman. However, the shortcomings of this approach have been exposed by Chappell. Social constructionism may be considered the main alternative, but this group of positions has also been the subject of some persuasive criticism, e.g., by Wiggins. Drawing on Millikan and Bach, I argue that a viable alternative would require a modification of certain ontological assumptions about the classification of real kinds and the specificity of socially constructed entities.

KEYWORDS

personhood, criterialism, social constructionism, historical kinds, essentialism

SŁOWA KLUCZOWE

osobowość, kryterializm, konstrukcjonizm społeczny, rodzaje historyczne, esencjalizm

1. INTRODUCTION

The debate about the relationship between a person's status and the category of rights to which he or she is entitled has been going on since at least the early modern period.¹ Although in the modern philosophy of law these issues are often analysed independently, they are still closely related areas, largely because of the practical nature of the questions surrounding them: to whom do the rights belong, to what extent, and on what basis? It seems, therefore, that the reflection on the concept of subjective rights cannot do without some axiological and ontological considerations regarding the category of persons: the holders of rights. The purpose of this paper is, therefore, to reflect on the consequences for legal theory of the critique of criterialism that dominates the modern philosophy of personhood: the assumption that the classification of entities that can be characterized as persons follows from their fulfilment of certain necessary and sufficient conditions. This assumption stems from a series of questions about the nature of personhood that have been analysed at least since the time of John Locke (1690), and more recently have found expression in the work of authors such as Harry Frankfurt (1971), Daniel Dennett (1976), Peter A. French (1983), or Michael F. Goodman (1992). The drawbacks of theories based on this assumption have been pointed out by Timothy Chappell, among others, but the question of the possibility of formulating a convincing alternative remains open. In this paper, I will, therefore, try to outline some possible developments of philosophical positions, with particular reference to the conceptual framework proposed by Ruth G. Millikan, and assess their relevance to the theoretical-legal issues at stake.

Given the standard image of the legal domain as consisting of either things or persons, it is safe to say that the question of who can hold rights is essentially the same as the question of the nature of legal personhood. Until recently, the topic of legal personhood has been most extensively discussed in relation to corporate entities, although it should be noted that in recent years matters relating to the personhood of animals or 'environmental persons' have been explored increas-

¹ See, e.g., Visa AJ Kurki, *A Theory of Legal Personhood*, Oxford University Press, Oxford 2019.

ingly, and the status of artificial intelligence has become a burning issue.² Nonetheless, it seems relatively rare to find in the literature a more holistic perspective that goes beyond the particularistic questions of strictly theoretical-legal, possibly ethical issues. One such position, however, is Ngaire Naffine's *Law's Meaning of Life*, in which the author proposes a useful conceptual apparatus involving a division of positions in the philosophy of law. According to Naffine, a group of views referred to as 'Realist' makes the legal notion of personhood dependent on a broader (philosophical or even commonsensical) account, whereas 'Legalists' argue that these approaches are (and should be) independent. Realists (more precisely, 'Metaphysical Realists') therefore place certain preconditions for something to be legally recognized as a person. Naffine indicates that what these preconditions are depends on which subgroup of realists are we talking about, for realists can be either 'Rationalists', 'Religionists' or 'Naturalists'.³ In this paper, however, without necessarily disagreeing with Naffine, I would like to explore a slightly different approach. Namely, that the problem with the ontological consideration of personhood may not be the content of the concept of a person itself, but rather the method of establishing it, and certain assumptions about how we actually go about doing so. Moreover, unlike Naffine, I will not pay much attention to legal theory, but rather focus on positions that I tend to characterize as primarily philosophical (i.e. formulated independently of strictly theoretical-legal issues).⁴ I will, therefore, concentrate on analysing the implications of adopting the dominant approach in contemporary analytic philosophy: criterialism. If the argument I present is valid, it should also work in favour of the thesis that personhood constitutes a 'historical' kind, rather than a 'traditional' natural kind (see Section 4).

Referring to P. F. Strawson's terminology, it can be said that the purpose of this work will be a study in descriptive metaphysics: an attempt to characterize the ontological assumptions that function at the basis of our conceptual apparatus and, more or less explicitly, determine our axiology.⁵ At the same time, it should be emphasized that it seems particularly valuable to study such philosophical

² See Ngaire Naffine, *Law's Meaning of Life: Philosophy, Religion, Darwin and the Legal Person*, Hart, Oxford 2009, 11, 16; and Visa AJ Kurki, Tomasz Pietrzykowski (eds), *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, Springer, Cham 2017.

³ Ngaire Naffine, *Law's Meaning of Life*, 20–24.

⁴ For instance, it is possible to point to several theories of legal personhood that are based specifically on the aforementioned question of the status of corporations. It should be noted, however, that these theories often do not answer the question of the general nature of personhood but rather focus on its source. This does not mean that they cannot be classified according to Naffine's terminology. For example, under concession theory, legal personhood is a status conferred by the State to promote economic or social goals. It is, therefore, conditional and purpose-oriented, placing it within the broadly understood 'Legalist' category.

⁵ Peter F Strawson, *Individuals. An Essay in Descriptive Metaphysics*, Routledge, London and New York 2003.

positions in which axiology has not prejudged the answer to all the questions at the start of the inquiry, for example, by theological justifications. It is also worth emphasizing that there are significant differences in the understanding of concepts between different areas of study, such as general philosophy and law. For instance, moral philosophers or metaphysicians often disagree whether non-human individuals of some sort can be classified as persons; however, as we already know, legal theory assumes that this is possible and, in fact, recognizes many such entities. Sometimes it looks downright anecdotal in the literature. At the same time, one author can point to a list of criteria that an entity must meet to be considered a person, such as the ability to think and the self-awareness of being a person. On the other hand, the same author can say without much trouble, to quote Mary Midgley, that: ‘The law then can, if it chooses, create persons; it is not merely a passive recorder of their presence [...]’.⁶ So, can the law actually create thinking and self-aware entities? It sounds quite caricatured, but it stems from the long-standing confusion of descriptive and normative approaches, the latter of course being picked up most often by legal theorists. To recapitulate, without looking at the concept of a person in a broader context, we cannot fully answer the two key questions in this debate: what do realists actually claim and what philosophically interesting consequences does legalism generate?

The structure of this paper is, therefore, as follows. First, I will outline criterialism as the dominant methodological approach in contemporary philosophical thinking about persons. Then I will point out why criterialism fails, with emphasis on Timothy Chappell’s critique. Second, I will briefly describe radical constructionism as an alternative approach. In the same vein, I will discuss why radical constructionism does not seem to work either. Here, David Wiggins’s critique will be especially helpful. Third, I will consider what follows from that for the philosophy of law, and whether we have a viable alternative to these positions. Finally, I will try to show an approach that relates to the conceptual apparatus of social ontology, which I refer to as ‘deflationist constructionism’, and the category of ‘historical kinds’, as well as indicate their relation to the methodological issues that emerge in the debate.

2. CRITERIALISM AND WHY IT FAILS

To begin with, it should be noted that the issues concerning ontology of a person can be divided into two main categories: the definitional matters and the questions related to personal identity over time. John Locke was one of the first

⁶ Mary Midgley, *Persons and Non-Persons*, in Peter Singer (ed), *In Defense of Animals*, Harper Perennial, New York 1986, 54.

to explicitly mark this division and postulate the separation of these two areas of inquiry, stating in his 1690's *Essay Concerning Human Understanding* that 'to find wherein *personal Identity* consists, we must consider what *Person* stands for'.⁷ Quite repeatedly, however, the discussion of the concept of a person in philosophical, and increasingly often also theoretical-legal, contexts begins with the question of the criteria of diachronic identity of persons and, therefore, what makes a person at time t_1 still the same person at time t_2 . I leave this problem aside because it seems that a good part of the discussion starts from intuitive assumptions, but does not tell us what a person is at all (some major authors writing about the persistence over time of persons, such as Eric Olson, explicitly state that the question of 'what a person is' is of no interest to them).⁸ On the one hand, diachronic analyses are crucial from the perspective of questions such as whether a human person is identical to a given human organism (i.e., an animal) but, on the other hand, it seems that one should first follow Locke and, especially in the philosophical-legal perspective, take a step back and consider the concept of a person in static, definitional terms. Locke himself proposes the following 'static' definition: '[a person] is a thinking intelligent Being, that has reason and reflection, and can consider it self as itself, the same thinking thing in different times and places [...]'.⁹ It is clear from this passage that Locke formulates the definition by stating the necessary and sufficient conditions for being a person. Since at the same time these conditions constitute the criteria for assessing whether an object belongs to the extension of the predicate 'is a person', Locke thus seems to establish the criterialist approach to the problem of personhood. The list of other examples of significant criterialist accounts includes the theories of at least several prominent contemporary thinkers. The conditions of being a person, according to the following authors, are:

- Harry Frankfurt: occurrence of second-order acts of will.¹⁰
- Daniel Dennett: rationality, the ascription of intentional predicates, the possibility of adopting a 'personhood attitude' towards such a being and its ability to reciprocate, the ability to communicate verbally, and self-consciousness.¹¹

⁷ John Locke, *An Essay Concerning Human Understanding*, Peter H Nidditch (ed), OUP, Oxford 1979, 335.

⁸ Eric T Olson, *What Are We?: A Study in Personal Ontology*, Oxford University Press, Oxford 2007, 44. See also EJ Lowe, who stresses that: 'We cannot hope to formulate an adequate criterion of identity for objects of a kind *K* unless we have an adequate idea of what *Ks are*'. EJ Lowe, *The Probable Simplicity of Personal Identity*, in Georg Gasser, Matthias Stefan (eds), *Personal Identity: Complex or Simple?*, Cambridge University Press, Cambridge 2012, 143.

⁹ John Locke, *An Essay*, 335.

¹⁰ Harry G Frankfurt, 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 1971, 68 (1), 5–20.

¹¹ Daniel Dennett, *Conditions of Personhood*, in Amélie Oksenberg Rorty (ed), *The Identities of Persons*, University of California Press, Berkeley 1976, 177–179. Dennett claims that the

- Peter A. French: manifestation of behaviour that can be considered intentional (in Davidsonian terms).¹²
- Michael F. Goodman: the ability to have moral consciousness.¹³

Again, I should point out that I am restricting myself to this particular (though dominant in contemporary philosophy) methodological tradition, which I have called ‘Lockean’ because I want to start from the question of the ontology of a person, rather than from axiological considerations. The concept of a person is obviously crucial, for example, from the point of view of Christian philosophy, but it is strongly entangled with theological or eschatological assumptions, and much of the analysis has a fundamentally ethical dimension, starting from the notion of the inherent dignity of a human being. For this reason, I omit a discussion of, for example, most personalist accounts, although by definition they are related to the question at hand.¹⁴

Having established the outlines of the dominant methodological approach to the formation of the concept of a person, we can move on to the first fundamental question. Does the criterialist approach actually work? The answer, which I will try to justify later in the text, is twofold, and the second part will be of particular interest: in the context of applying law – yes. In philosophical and even colloquial contexts, not necessarily. Why is that so? Because in the problem under discussion, we are not dealing with legal interpretation (where we already have a definition and check whether something belongs to the extension of a given predicate), but we must take a step back to determine precisely why we treat any kind of being as a person. There are, however, some fundamental problems with the criterialist approach in this regard. The first objection is self-evident: it is the arbitrariness and non-intuitiveness of the lists of criteria. Very often, elaborate and ‘rigidized’ lists falsify the picture of our colloquial intuitions about the concept of the person, rendering them glaringly inadequate. This aspect of the criterialist approach was also acknowledged by Dennett, indeed one of its proponents, who admitted that ‘Human beings or other entities can only aspire to being approximations of the ideal, and there can be no way to set a “passing grade” that is not arbitrary’.¹⁵

The second objection, however, is more serious: do we ever conduct a ‘personhood test’ in practice? In a paper analysing criterialist approaches, Timothy Chappell argues that the opposite is true. We *first* recognize someone as a person, and only then do we expect them to manifest the typical characteristics of persons:

conditions listed are necessary, but ‘hard to say whether they are jointly sufficient conditions for moral personhood’, 177.

¹² Peter A French, ‘Kinds and Persons’, *Philosophy and Phenomenological Research* 1983, 44 (2), 251.

¹³ Michael F Goodman, ‘A Sufficient Condition for Personhood’, *The Personalist Forum* (Spring 1992 Supplement) 1992, 8 (1), 75–81.

¹⁴ It can be assumed that they would largely fall into the Naffine’s ‘Religionist’ category.

¹⁵ Daniel Dennett, *Conditions*, 193.

The criterialists are right to insist that properties like self-awareness, emotionality, rationality, and so forth, are crucial parts of our concept of a person. It does not follow that these properties can be used as the criterialists want to use them: as *tests* for personhood.¹⁶

Chappell argues that, considering this, we should rather regard the idea of a person as *proleptic*, i.e., anticipatory:

To treat someone as a person is to engage with him as the kind of creature to which that ideal applies. So to treat him is not, at the deepest level, a response to his *behavior* at all but to his nature. To see some creature as a person is to take an attitude toward that creature that, *before* any behavioral evidence comes in, is already different from our attitudes to creatures that (we think) are not persons.¹⁷

How exactly would this work? Here Chappell gives two rather convincing examples. The first one concerns raising children. Parents tend to treat children as persons from their birth, even though they are often fully aware of their actual level of psychological development – in other words, even though their child does not meet some of the often-indicated criteria for being a person. Chappell outlines roughly the following scenario: a young child has picked up a book, walks across the room and drops it in their sibling's lap. The parent's response would probably be: 'How nice of you to let them share your book!' Is that, however, what the child actually did? Chappell states that a perfectly legitimate response might be: 'Yes, if the parent reacts like that' – and *because* they react like that. The parent's reaction is the actual interpretation of the child's action. By means of a speech act, that possibly accidental behaviour has been given meaning and the replication of this interaction eventually allows the child *to obtain* personal qualities. Is this absurd? No, because at a certain stage of development, typically, but not necessarily exclusively, in the case of a human being, this will begin to translate into how the child will begin to behave intentionally. This is where this approach differs from our treatment of, for example, such non-human animals that typically do not reach this stage of moral and cognitive development:

For all he knew about it, his action might as well have been simply a random sequence of bodily movements. But now that his mother offers her interpretation, the child, on his mother's authority, learns to see himself a certain way. Because his mother frames his act as one with a certain meaning (as the giving of a gift), and because the child sees his mother as seeing his action this way, and because his mother sees the child as seeing her as seeing his action this way (Grice 1957), the action *comes to have* that meaning.¹⁸

¹⁶ Timothy Chappell, 'On the Very Idea of Criteria for Personhood', *Southern Journal of Philosophy* 2011, 49 (1), 12.

¹⁷ *ibid.*

¹⁸ *ibid.*, 9.

The second example, on the other hand, concerns the statistically low effectiveness of rehabilitation programmes for convicts. We know from experience that this type of criminal policy has little chance of success, but even if it seems that a particular criminal is not promising to improve, we still provide them with the opportunity to do so in principle, precisely because we have a certain idea of a subject who can change his behaviour.¹⁹

And getting back to the central issue of this paper, here comes a certain key generalization that can be described as Kantian, namely: no moral subject exemplifies *all the properties* of the ideal moral subject, but we treat others in light of this *idea* anyway. In other words: perhaps no person is an ideal person, meeting all the criteria on the list, but this does not mean that failure to meet the criteria excludes one from the circle of persons. The criteria are merely a secondary idealization:

In general, the idealization that accords X the status due to a ϕ -er does not have to involve any kind of expectation that X will ever actually ϕ , or even can ϕ . What it *does* have to involve is the understanding that X is of a kind that characteristically includes ϕ -ers.²⁰

It should be noted that Chappell himself conditionally admits the ‘personhood test’ proper to criterialists in extremely unusual, science-fictional situations, since he assumes that the paradigmatic case of a person is a human being, and that if we were dealing with a completely different entity, we might have to resort to such a procedure. However, as we shall see, because of our innate preconditioning when it comes to dealing with entities that manifest human characteristics, this kind of admission of criterialism is not so much unwarranted as probably unworkable.²¹

Since criterialism does not provide an adequate characterization of how the concept of a person works, do we have an alternative? It seems that at the opposite pole lies a group of anti-essentialist positions that can be described as radical constructionism. I use the term ‘radical’ here to indicate their difference from the more refined ‘deflationist’ constructionism, which I will discuss in a moment. I will confine myself here only to outlining Wiggins’s critique and his characterization of this group of positions, without going into much detail.

¹⁹ *ibid.*, 11.

²⁰ *ibid.*

²¹ *ibid.*, 19. See also Oswald Hanfling’s remarks on the topic. Hanfling, referring to Wittgenstein, suggests that even such conditional criterialism would not work (although he does not put it that way) because of the specificity of our attitudes (e.g., we would not, and should not, deny personhood to a robot looking and consistently behaving almost entirely like a human being, who is able to engage in certain ethical practices like promises, just because of his material constitution). Oswald Hanfling, ‘Machines as Persons?’, in David Cockburn (ed), *Human Beings*, Cambridge University Press, Cambridge 1991, 25–34.

3. RADICAL CONSTRUCTIONISM AND WHY IT FAILS

In the first edition of *Sameness and Substance* (1980), David Wiggins points out that the accounts that attempt to fix a systemic specification of a person by a finite number of attributes ‘without essential reference to human beings just as they are’ drastically impoverish our mode of interpretation of ‘speech, conduct and the thoughts that lie behind the desires that lie behind conduct’.²² Eventually, they may

license a state of affairs in which there was absolutely nothing except fear of confusion to obstruct proposals for modifying or reinventing even the *accepted* specification of what a person was – just as we constantly and effortlessly modify and refashion through time certain institution- and artifact-concepts.²³

David Bakhurst, commenting on Wiggins, expands on this idea, stating that if we somehow attempt to arbitrarily create the concept of a person by convention or even tacit agreement, i.e. ‘the marks of personhood are socially determined’, we are left with only the nominal essence of personhood, susceptible to change whenever ‘different patterns of use for the term “person” can be established’.²⁴ This results in a situation in which:

Hard cases about identity and survival, of a kind concocted in philosophical thought experiments, show the limits of our existing concepts, though nothing dictates how we should respond to them. [...] A robust conception of human nature, in contrast, enables a deeper politics that can admit that the compass of human desire, and the depth of human potential, may transcend our present conceptions.²⁵

Although Wiggins criticizes constructionism understood in this way (as a metaphysical position) primarily from moral positions (an indictment of the ‘managerial utilitarianism’ found in social engineering projects), his argument is primarily directed at the inadequacy of this approach. If the concept of a person fails to account for our moral practices and for certain key features of entities

²² David Wiggins, *Sameness and Substance*, Blackwell, Oxford 1980, 179. David Bakhurst notes that in *Sameness and Substance Renewed*, ‘Wiggins has cut the treatment of social constructionism in favour of an expanded discussion of neo-Lockean views’. David Bakhurst, ‘Wiggins on Persons and Human Nature’, *Philosophy and Phenomenological Research* 2005, 71 (2), 463. Space does not permit me to discuss the constructionist claims in detail, but, following Bakhurst, I would like to point out ‘that this kind of constructionism is not just a straw man’: see e.g., Derek Edwards, Michael Ashmore and Jonathan Potter, ‘Death and Furniture: the rhetoric politics and theology of bottom line arguments against relativism’, *History of the Human Sciences* 1995, 8 (2), 25–49, and Kenneth Gergen, *An Invitation to Social Constructionism* (London: Sage, 1999), 464, note 6.

²³ *ibid.*, 181.

²⁴ David Bakhurst, ‘Wiggins on Persons’, 463.

²⁵ *ibid.*

classified as persons (and Wiggins takes human beings as the paradigmatic case), then we cannot expect the concept to help us in our practical reasoning. Wiggins writes about certain capacities or a certain moral sensibility, and indeed it is possible to find examples in the literature where the concept of a person is linked to such capacities. For example, Charles Taylor writes about the capacity to feel shame or humiliation as occurring only among entities that we classify as persons.²⁶ To oversimplify somewhat the gist of Wiggins's critique: we cannot rule out the possibility that a failure to consider the specificity of human nature will backfire on us in practical endeavours.²⁷

There is also a secondary issue here. We can think of legalism as the philosophical-legal equivalent of radical constructionism. Is it, therefore, subject to the same objections? It would seem so, although as we already know, legalism does not explicitly refer to any philosophical positions and tends to avoid being categorized as one of them. However, the Wigginsian problem of adequacy can be easily translated into the functioning of the law in pragmatic terms, particularly its effectiveness. In other words, it is but another reminder that utter arbitrariness can easily be contrasted with Fullerian examples to demonstrate the dysfunctionality of laws completely detached from reality.

It seems that at this point of the inquiry we are faced with several questions. Do we have an alternative to criterialism ('traditional' essentialism) and radical constructionism (anti-essentialism)? Can realists successfully appeal to a non-legal notion of a person without exposing themselves to objections to criterialism? And eventually, can legalism convincingly answer the objections parallel to those raised against radical constructionism? The answer to all the above questions is yes, if we consider that the realist position can be somewhat weakened and certain limitations can be imposed on the legalist position. As we shall see, the tools for this can be provided by the conceptual apparatus of 'historical kinds', as well as some elucidations on what the construction of social entities consists of in causal and linguistic terms.

²⁶ Charles Taylor, *Human Agency and Language*, Cambridge University Press, Cambridge 1985, 102–111.

²⁷ See also Daniel Robinson, who points out that, for both cognitive and practical reasons, we cannot afford to embrace radical relativism about personhood: the concept only works in our practical reasonings if it is relatively stable. Daniel N Robinson, 'Personhood: What's in a Name?', *New Ideas in Psychology* 2012, 30 (1), 93; and Christine Korsgaard's critique of Derek Parfit's take on personhood (as 'a matter of mere grammatical convenience'), which also focuses on practical reasoning. Christine M Korsgaard, 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', *Philosophy & Public Affairs* 1989, 18 (2), 131–132.

4. DO WE HAVE AN ALTERNATIVE? DEFLATIONIST CONSTRUCTIONISM AND HISTORICAL KINDS

Before we go any further, it should be noted, in the context of our critique of criterialism, that the fact that we cannot give necessary and sufficient criteria for something does not mean that it does not exist. This is convincingly emphasized by Patricia Churchland when she criticizes Farah and Heberlein for drawing far-fetched conclusions from their analysis of neuroscientific, empirical attempts to define personhood.²⁸ In their 2007 paper, Farah and Heberlein explore the possibility of there being a ‘natural kind’ in the world that corresponds to persons, of which necessary and sufficient properties (‘objective and clear-cut biological criteria’ corresponding with common intuitions) would be discovered by collecting empirical data in order to develop ‘a scientific taxonomy in place of a folk taxonomy’.²⁹ Farah and Heberlein claim that although it may be possible for this approach to replace more traditional ones, based on psychological notions, it will face similar problems related to the lack of clear category boundaries.³⁰ Finally, they conclude that personhood, however useful as a concept, does not delimit a natural kind, and is merely an illusion.³¹ On the contrary, referring to the work of Eleanor Rosch, who in the 1970s showed ‘that most everyday working categories typically have a radial structure’, Churchland states that it would be absurd to claim, for example, that diseases do not exist simply because we have problems with necessary and sufficient conditions for membership of the category due to the multiplicity of borderline cases.³² However, this is not in itself a significant problem for criterialism. After all, we have a variety of categories whose application causes similar difficulties. But if we are in favour of traditional essentialism about real kinds (in which case ‘real’ would stand for ‘natural’), then, following Brian Ellis, a proponent of ‘scientific essentialism’, the less clear the boundaries of the categories, the less chance there is that we can actually talk about essences in a given domain:

As we move to yet more complex systems, from biological organisms up to ecological or social systems, natural kinds analyses become much less interesting. There are no natural kinds that satisfy the strict criteria applicable to chemical kinds that can readily be distinguished, and there are no sets of intrinsic characteristics of ecological, economic, social or other high-level systems that could plausibly be used to

²⁸ Patricia S Churchland, ‘The Necessary-And-Sufficient Boondoggle’, *American Journal of Bioethics* 2007, 7 (1), 54–55.

²⁹ Martha J Farah, Andrea S. Heberlein, ‘Personhood and Neuroscience: Naturalizing or Nihilating?’, *American Journal of Bioethics* 2007, 7 (1), 39.

³⁰ *ibid.*, 40.

³¹ *ibid.*, 45–46.

³² Patricia S Churchland, ‘The Necessary-And-Sufficient Boondoggle’, 54.

define appropriate microspecies (as genetic make-ups might be used to define biological microspecies).³³

This, together with the fact that the lists of criteria are extremely diverse and, as already mentioned, give the impression of being arbitrary, may mean that we are dealing with entities that are rather part of social reality, although not in the sense that a legalist or radical constructivist would postulate. As Adina Roskies points out:

It is not the case that the only concepts that can be reified are concepts that correspond to natural kinds. Natural kinds are not the only kinds we care about, nor are they the only things that exist. [...] For certainly there *are* persons, reified perhaps not by necessary and sufficient physical criteria, but rather by our social structure, our customs, our intuitions, and our history, as well as our biology. Persons may be social/ biological constructs of a sort, but they are not illusory for that.³⁴

That even natural kinds depend strictly on social and cultural conventions is, of course, nothing new. Writing about Lockean nominal essences, the ideas that allow us, for example, to group similar objects into classes, Daniel Robinson indicates that: ‘These characterizations arise from conventional discourse, the contingencies of culture and context, the nuances of perception, memory and mental life. Locke’s nominal essence, in a word, is *constructed*’.³⁵ Robinson refers here to Alexander Bird’s example of various approaches to vegetables: at a greengrocer’s we can be sure to encounter different classifications of such plants than the strictly biological ones.³⁶ But if, even at such a seemingly basic level, we encounter constant discrepancies between classifications, whether past and present or even simultaneous, can we even speak of the existence of kinds that are anything other than the product of our more or less arbitrary choices? It seems so, and as we shall see, the problems associated with traditional essentialism do not necessarily imply a capitulation to radical constructionism. What I will try to do here is to outline a proposal for a certain way of thinking about socially constructed entities that seems acceptable to both realists and legalists. Then, in the remainder of this section, I will go on to explain how this approach can provide a background for an alternative to traditional essentialist realism.

What I would like to call ‘deflationist constructionism’ does not actually constitute a separate theory. Rather, it is a way of thinking about the conceptual apparatus of social ontology devoid of strong ontological commitments and epistemological assumptions about the nature of socially created entities. I refer here

³³ Brian Ellis, *The Philosophy of Nature: A Guide to the New Essentialism*, Acumen, Chesham 2002, 32.

³⁴ Adina L Roskies, ‘The Illusion of Personhood’, *American Journal of Bioethics* 2007, 7 (1), 56.

³⁵ Daniel Robinson, ‘Personhood’, 91.

³⁶ *ibid.*, 92.

to Ruth G. Millikan's short commentary on this very topic, in which she indicates how one can think about the role of two key concepts in this issue: intention and convention.

First, Millikan indicates that we need to be careful about the way we perceive the construction of social beings by intention, and that misunderstandings in this area arise from the confusion of two different, perfectly ordinary phenomena. The term 'socially constructed', as applied to such kinds as women or Brahmins, does not actually imply that there is something special about the ontology here:

They are on the one hand *socially caused* and on the other they are *socially named*, and it is best not to run these two ideas together. That traditional women and Brahmins are members of their respective kinds is *not*, say, 'merely relative to the intentionality of agents'. Thinking may have made them members of kinds, but by *causing* them to be as they are, putting them under certain names, but not, of course, by fiat.³⁷

Second, Millikan also makes it clear that our thinking about women or Brahmins is not unlike our thinking about *other* real kinds, while at the same time emphasizing that we do not need to know the mechanisms of selection, to which we will return shortly, that led to their creation, 'clustering their various features together'.³⁸ By analogy, we usually do not know them in the various animal species, for example, and this in no way prevents us from thinking about rabbits in a perfectly ordinary way.³⁹ In other words, when it comes to a socially caused and named kind, we do not have to 'understand the origin of that kind as social but merely [...] grasp its current character'.⁴⁰ To sum up this point in Millikan's own words:

What has sometimes been called 'social *constitution*' here is merely a matter (1) of having a *causal* social history, a history of social reproduction, and (2) that history being semantically required for application of a thing's name. [...] Nothing of a new ontological kind has entered the world here. Copying certain patterns on purpose and giving names to these patterns and to some of their parts – that is all that has occurred.⁴¹

We can now return to the question of natural kinds. Since it seems that the realist needs a strong notion of person, which refers to something 'out there in the world', and since all the evidence so far suggests that it is difficult to treat person-

³⁷ Ruth G Millikan, *Deflating Socially Constructed Objects: What Thoughts Do to the World*, in Mattia Gallotti, John Michael (eds), *Perspectives on Social Ontology and Social Cognition*, Springer, Dordrecht 2014, 28.

³⁸ *ibid.*, 29.

³⁹ For a slightly different argument for this claim, see Francesco Guala's critique of 'the difference thesis' concerning natural and institutional kinds. Francesco Guala, *On the Nature of Social Kinds*, in Mattia Gallotti, John Michael (eds), *Perspectives on Social Ontology and Social Cognition*, Springer, Dordrecht 2014, 57–68.

⁴⁰ Ruth G Millikan, *Deflating*, 29.

⁴¹ *ibid.*, 33.

hood as a natural kind, as it is heavily entangled in cultural and social context, can we somehow defend this position? Here, again following Millikan, it is possible to divide real kinds in such a way as to break out of this entanglement.

In her 1999 paper, Millikan introduces the notion of *historical kinds*.⁴² Unlike an *eternal kind*, which is ‘held together by universal and eternal laws of nature’ (recall Ellis’s comment on the ‘strictness’ of chemical kinds), a *historical kind* is formed because ‘there is a certain historical link between the members of the kind that causes the members to be like one another’.⁴³ Therefore, on the one hand, we have kinds that exist because of certain intrinsic properties of their members, for which ‘spatiotemporal locations play no role in the explanation of their similarities’.⁴⁴ On the other hand, we have clusters created because ‘something akin to reproduction or copying has produced all the various kind members from one another or from the same models’ and ‘the various kind members have been produced in or in response to the very same ongoing historical environment’.⁴⁵ Additionally, it is usually also true that ‘some “function” is served by members of the kind, where function is understood roughly in the biological sense as an effect raising the probability that its cause will be reproduced [...]’.⁴⁶ It is important to note that replication and selection are universal processes, not limited to the field of biology (think of how a child can repeat its actions to win praise from its parents, as in Chappell’s example).⁴⁷ Needless to say, ‘historical kinds are unlikely to ground exceptionless generalizations’, but they still are ‘domains over which predicates are non-accidentally projectable’, allowing for successful inductions.⁴⁸

What is crucial, as shown for example by Theodore Bach, this type of analysis allows for a good characterization of social kinds. Drawing on Millikan, Bach states that gender kinds are ‘natural kinds with a historical essence and also that members of a gender kind possess a common teleological function’.⁴⁹ Although we are still dealing here with the category of ‘essence’, it is precisely its *historical*, not its *eternal*, variety that is at issue. Bach claims that this ‘explanatory approach [to natural kinds] frees the notion of kind essence from its traditional and problematic implications. The essence [...] on this view is the source that organizes,

⁴² Ruth G Millikan, ‘Historical Kinds and the “Special Sciences”’, *Philosophical Studies* 1999, 95 (1), 45–65. More recently, Millikan elaborated on this notion in her later books, such as *Language: A Biological Model* (OUP, Oxford 2005) and *Beyond Concepts* (OUP, Oxford 2017).

⁴³ Ruth G Millikan, ‘Historical Kinds’, 51, 55.

⁴⁴ Theodore Bach, ‘Gender Is a Natural Kind with a Historical Essence’, *Ethics* 2012, 122 (2), 242.

⁴⁵ Ruth G Millikan, ‘Historical Kinds’, 55.

⁴⁶ Ruth G Millikan, *On Clear and Confused Ideas: An Essay About Substance Concepts*, Cambridge University Press, Cambridge 2000, 20.

⁴⁷ Theodore Bach, *Gender*, 246.

⁴⁸ Ruth G Millikan, ‘Historical Kinds’, 55.

⁴⁹ Theodore Bach, *Gender*, 232.

or accounts for, reliable property correlations'.⁵⁰ According to Bach, 'this appeal to essence and function achieves a better solution to the theoretical and political problems that motivate anti-essentialist accounts of gender'.⁵¹ Since Bach's analysis shows exactly how a social kind can be characterized without falling into traditional essentialism, on the one hand, and radical constructionism, on the other, this approach seems perfectly apt for the notion of personhood as well:

If genders are historical kinds, then being a woman or a man is not a matter of possessing some set of biological or psychological properties, as these are merely probabilistic indicators of a deeper historical unity. Nor, for the same reason, is it to occupy a certain social position [...]. Being a woman or a man, I now argue, is to have the right sort of origin and replicative history in relation to a more fundamental historical kind – a replicating gender system.⁵²

If we take an analogous view of the concept of personhood, we can see how our default 'personhood stance' towards human beings can be traced back to certain biological features to which we have been conditioned by years of evolution.⁵³ At the same time, it shows how the class of persons can be expanded in a non-arbitrary way, if only we pay attention to how it has been historically shaped in light of the function the concept plays in our social reality. Simultaneously, we can see how criterialism proves to be too superficial. For, on the one hand, it is a surprisingly liberal position, allowing for a classification of potentially diverse entities as persons. On the other hand, it does so in a way that does not correspond to how the concept of personhood works in the first place. Now let us revisit the constructivist (legalist) strategy of defining personhood and compare it to the one based on the notion of historical kinds. Bach points to the existence of an analogous approach in the debate on gender, which involves a recourse to the notion of the objective type, which is a non-discriminating, mind-independent similarity class:

Objective types include 'items on a desk', 'items outside the moon's orbit', and 'all the atoms that are not presently in the Queen of England's left thumb nail'. In addi-

⁵⁰ *ibid*, 241.

⁵¹ *ibid*, 232.

⁵² *ibid*, 246.

⁵³ 'Why would such a misleading system for person representation have evolved? The answer most likely concerns the intensely social nature of our species and perhaps the rarity of ambiguous cases of personhood in our evolutionary history. Like other social species, our individual survival depends on relating successfully to our conspecifics. More for us than for other species, this requires understanding the immensely complex behaviors that result from their beliefs, motivations, and personalities. As the anthropologist Guthrie (1995) has observed, in discussing religious belief systems, the cost of attributing intentionality to some non-intentional systems may be less than the cost of failing to adopt the intentional stance toward some systems that are intentional. In other words, it may have been adaptive to err on the side of activating the personhood network too often'. Farah and Heberlein, 'Personhood', 45. It is worth noting that Chappell also assumes that the pre-identification of potential persons demonstrated in his examples is based on biological features.

tion to being nondiscriminatory, objective types are nonexplanatory: claiming that a group of items is an objective type is to state a unity rather than explain a unity.⁵⁴

But in the context of what we already know about arbitrary classifications, it is easy to see the flaws in such a strategy. If there are no membership conditions other than ad hoc pragmatic considerations, the notion of personhood loses both its explanatory (recall the ‘explanatory approach’ to real kinds) and, possibly, symbolic function. In this case, however, how can we even speak of extending the category of persons beyond, say, the paradigmatic cases of human beings? Again, we are not dealing with the simple ‘construction’ of a new category, for as Bach states, ‘an ontology of historical kinds provides explanations for objective similarities’.⁵⁵ Bach, following Millikan, points out that ‘The empirical processes that ground kind membership for historical kinds are often uneven’, and ‘in these cases, the resulting historical kind will have vague boundaries’, unlike eternal kinds, steadily grounded by their intrinsic properties.⁵⁶ Furthermore, Bach adds that ‘transitions between historical kinds are also vague’ because ‘distinct ontogenetic processes displace older ones’.⁵⁷ Put in this way, the range of entities we can classify as persons changes significantly, and for a variety of reasons. For the fundamental question is one of historical connections, not similarity.⁵⁸ This means that, given one’s ontogeny, failure to meet the criteria of personhood does not mean that one ceases to belong to the real kind of persons, just as in the case of gender according to Bach:

By identifying womanhood with historical properties, historical essentialism does not provide a standard of occurrent behavior or social positioning that individuals must approximate in order to achieve authentic status as a woman.⁵⁹

Nonetheless, further evolution and branching of the lineage of persons can take place in isolation from the original model, so that although the default category of personhood is grounded in our neurobiology and our relationship to other human entities, it is conceivable that an extension of it could occur without any direct analogy, such as in the case of environmental entities. Essentially, this means that the real historical kind of persons is not foreclosed by arbitrary normative considerations, but can evolve according to how the lineages that constitute the concept of personhood account for its function in our psychological, moral or legal discourses. At the same time, the stability of a historical kind is

⁵⁴ Theodore Bach, ‘Gender’, 271.

⁵⁵ *ibid.*

⁵⁶ *ibid.*, 269.

⁵⁷ *ibid.*

⁵⁸ *ibid.*, 264.

⁵⁹ *ibid.*, 265. Similarly, Joseph Vukov argues that, in the context of the moral status of human beings, it is not the actual cognitive capacities that are decisive, but merely whether one belongs to the natural kind ‘human being’. Joseph Vukov, ‘Personhood and Natural Kinds: Why Cognitive Status Need Not Affect Moral Status’, *Journal of Medicine and Philosophy* 2017, 42 (3), 262.

determined by its function, and the fact that some tokens do not perform a typical role despite proper ontogeny does not mean that the kind itself is not clustered for some specific historical reason. But what would be the processes of replication that produce the historical kind of persons? Returning to deflationist constructionism and the question of convention, much like in Chappell's example of a child and its parent, this is largely done in language: 'hearers continue to respond in conventional ways, for example, by believing or by doing what they are told, because, often enough, the result is rewarding for them'.⁶⁰ There is, of course, more to the story than the mere reproduction of linguistic behaviour. It is important to note that the analysis of the notion of personhood in terms of historical kinds is not incompatible with the partial conclusions of Farah and Heberlein's paper: 'The real contribution of neuroscience to understanding personhood may be in revealing not what persons are, but rather why we have the intuition that there are persons'.⁶¹ The consequences of our neurobiological conditioning, which causes us to tend to regard persons as something different from the rest of the environment, are not only in no way illusory, but are a key part of the process of producing the original components of the concept of a person, and thus the model whose replication led to the development of the lineage. It is only by starting from paradigmatic cases grounded in psychology (or neuroscience) and language that we can extend our 'personhood stance' to other kinds of entities, but the concept of a person is not the product of an arbitrary list of criteria, and if we do not have tests for personhood, it is precisely because persons are constructed in this 'cautious' sense. As mentioned in the context of Wiggins's critique, overlooking this genesis and development can damage the relevance of the notion of personhood in our practical reasoning.

5. CONCLUSIONS

To conclude, we can now return to the original set of theoretical positions and make a brief assessment of their chances of success in defining the concept of a person. Since it seems quite impossible to find a satisfactory 'traditionally' essentialist account of personhood (as an extra-legal point of reference), realism as a position based on methodological criterialism seems doomed to failure. Thus, from a methodological point of view, legalism would seem to have the upper hand. However, as the critique of 'radical constructionism' suggests, here too the indictment of arbitrariness may prove fatal to the theory, as it does not provide convincing arguments for why it uses the concept of a person in the way

⁶⁰ Millikan, *Deflating*, 34.

⁶¹ Farah and Heberlein, 'Personhood', 40.

it does either, for example by giving the same status to both human beings and very different objects.⁶² In contrast, it seems that a more nuanced view of social constructionism can provide a convincing alternative. On the one hand, it allows us to see that the concept of personhood is indeed inherently dependent on the social reality. On the other hand, it is compatible with a view of its genesis that does not deny the ‘realness’ of personhood as a kind, but situates it within an improved classification – namely as one with a historical essence. The approach based on historical kinds accounts, for instance, for the fact that we are beginning to acknowledge the possibility of AI personhood not only because it possesses certain cognitive abilities or autonomy but also due to social and technological contexts that necessitate new legal frameworks. As AI becomes indispensable to society, its personhood may emerge as a practical solution to legal and governance challenges, with the criteria defining the threshold of human-like qualities serving as just one indicator of the viability of such a process. Similarly, environmental personhood has developed in response to shifting societal values and environmental crises, but can also be traced back to how local cultures perceive such entities. Such an approach seems to be a satisfactory point of reference for realists, and for legalists it provides tools for refining their constructivist endeavours. This does not mean, however, that we should not continue to search for further convincing methodological refinements that will allow us to better understand what personhood is. Perhaps there are equally convincing alternatives in other philosophical currents, or the present definitional inquiry can be fruitfully supplemented by theories of personal identity over time: the question of the nature of personhood remains rather open.

The work was supported by the National Science Centre, Poland, under research project no [2020/36/C/HS5/00600], titled *Analysis of the Concept of a Legal Person from an Ontological and Linguistic Perspective*.

REFERENCES

- Bach T, ‘Gender Is a Natural Kind with a Historical Essence’, *Ethics* (2012), 122 (2), 231–272
- Bakhurst D, ‘Wiggins on Persons and Human Nature’, *Philosophy and Phenomenological Research* (2005), 71 (2), 462–69
- Chappell T, ‘On the Very Idea of Criteria for Personhood’, *Southern Journal of Philosophy* (2011), 49 (1), 1–27

⁶² See Naffine’s comments on Nekam. Naffine, *Law’s Meaning of Life*, 178.

- Churchland PS, 'The Necessary-And-Sufficient Boondoggle', *American Journal of Bioethics* (2007), 7 (1), 54–55
- Dennett D, *Conditions of Personhood*, in Oksenberg Rorty A (ed), *The Identities of Persons*, University of California Press, Berkeley 1976, 175–196
- Ellis B, *The Philosophy of Nature: A Guide to the New Essentialism*, Acumen, Chesham 2002
- Farah MJ, Heberlein AS, 'Personhood and Neuroscience: Naturalizing or Nihilating?', *The American Journal of Bioethics* 2007, 7 (1), 37–48
- Frankfurt HG, 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy* 1971, 68 (1), 5–20
- French PA, 'Kinds and Persons', *Philosophy and Phenomenological Research* 1983, 44 (2), 241–25
- Goodman MF, 'A Sufficient Condition for Personhood', *The Personalist Forum* (Spring 1992 Supplement) 1992, 8 (1), 75–81
- Guala F, *On the Nature of Social Kinds*, in Gallotti M, Michael J (eds), *Perspectives on Social Ontology and Social Cognition*, Springer, Dordrecht 2014, 57–68
- Hanfling O, 'Machines as Persons?', in Cockburn D (ed), *Human Beings*, Cambridge University Press, Cambridge 1991, 25–34
- Korsgaard CM, 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', *Philosophy & Public Affairs* 1989, 18 (2), 101–132
- Kurki VAJ, Pietrzykowski T (eds), *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, Springer, Cham 2017
- Kurki VAJ, *A Theory of Legal Personhood*, Oxford University Press, Oxford 2019
- Locke J, *An Essay Concerning Human Understanding*, Nidditch PH (ed), OUP, Oxford 1979
- Lowe EJ, *The Probable Simplicity of Personal Identity*, in Gasser G, Stefan M (eds), *Personal Identity: Complex or Simple?*, Cambridge University Press, Cambridge 2012, 137–155
- Midgley M, *Persons and Non-Persons*, in Singer P (ed), *In Defense of Animals*, Harper Perennial, New York 1986, 52–62
- Millikan RG, *Deflating Socially Constructed Objects: What Thoughts Do to the World*, in Gallotti M, Michael J (eds), *Perspectives on Social Ontology and Social Cognition*, Springer, Dordrecht 2014, 27–39
- –, 'Historical Kinds and the "Special Sciences"', *Philosophical Studies* 1999, 95 (1), 45–65
- –, *On Clear and Confused Ideas: An Essay About Substance Concepts*, Cambridge University Press, Cambridge 2000
- Naffine N, *Law's Meaning of Life: Philosophy, Religion, Darwin and the Legal Person*, Hart, Oxford 2009
- Olson ET, *What Are We?: A Study in Personal Ontology*, Oxford University Press, Oxford 2007
- Robinson DN, 'Personhood: What's in a Name?', *New Ideas in Psychology* 2012, 30 (1), 89–96
- Roskies AL, 'The Illusion of Personhood', *The American Journal of Bioethics* 2007, 7 (1), 55–57

Strawson PF, *Individuals. An Essay in Descriptive Metaphysics*, Routledge, London and New York 2003

Taylor C, *Human Agency and Language*, Cambridge University Press, Cambridge 1985

Vukov J, 'Personhood and Natural Kinds: Why Cognitive Status Need Not Affect Moral Status', *The Journal of Medicine and Philosophy* 2017, 42 (3), 261–277

Wiggins D, *Sameness and Substance*, Blackwell, Oxford 1980