Economics & Philosophy (2024), **40**, 310–336
doi:10.1017/S0266267123000123

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Identity, ethics and behavioural welfare economics

Ivan Mitrouchev[1,*] and Valerio Buonomo[2]

[1]IESEG School of Management, Univ. Lille, CNRS, UMR 9221 – LEM – Lille Economie Management, F-59000 Lille, France and [2]Centre for Philosophy of Time, Department of Philosophy, Università degli Studi di Milano, Milan, Italy
*Corresponding author: Ivan Mitrouchev; Email: i.mitrouchev@ieseg.fr

(Received 2 July 2020; revised 11 January 2023; accepted 31 January 2023; first published online 27 April 2023)

Abstract
Multiple selves is a conventional assumption in behavioural welfare economics for modelling intrapersonal well-being. Yet an important question is which self has normative authority over others. In this paper, we advance an argument for what we call the 'ontological approach' to personal identity in behavioural welfare economics. According to this approach, ethical questions – such as which preference should be granted normative authority over another – can be informed by the ontological criterion of personal persistence, which aims at determining what it takes for an individual to persist from one time to another.

Keywords: Multiple selves; ontology; preference reversals; time; well-being

JEL codes: B41; D15; I31

1. Introduction

With the growing interest of behavioural economics in the evaluation, recommendation and prescription of public policy (Thaler and Sunstein 2003, 2009), a conventional assumption is to consider each individual as being composed of at least two selves: a far-sighted 'planner' and a myopic 'doer' (Thaler and Shefrin 1981), 'hot' and 'cold' states (Camerer *et al.* 2003), or an automatic system 1 and a reflective system 2 (Kahneman 2011). In a nutshell, the myopic doer/hot states/system 1 are the selves in which decisions are driven by fast thinking or made in the heat of the moment (e.g. eating a cake), whereas the far-sighted planner/cold states/system 2 are the selves in which decisions are driven by slow thinking or

© The Author(s), 2023. Published by Cambridge University Press.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

made reasonably (e.g. avoiding the temptation of eating a cake).[1] This dual conception assumes that individuals may take decisions they later regret, and that their normative authority is better located in the far-sighted planner, cold states or system 2.[2]

However, such an assumption of multiple selves presents a major difficulty from a philosophical viewpoint: how to locate individual normative authority when it is left unclear which of the preferences of the many possible selves are truly normatively relevant? In an appraisal of the value of individual autonomy in libertarian paternalism, Sunstein (2019) acknowledges this problem when he raises doubts regarding the arbitrariness of making ethical judgements about which self has normative authority over the other. In his words,

> What doers do might be one of the most significant and best experiences of their lives – even if they would have chosen otherwise in advance and perhaps even if they regret it afterwards. ... Why does John or Edith deserve authority at Time 1 or Time 3, rather than Time 2? What makes either of their views authoritative or authentic, rather than the choice at Time 2? (Sunstein 2019: 69–75)

Similarly, Kahneman (1994) raises an important ethical concern when observing the conflicting evaluations of patients during and after being subject to painful experiments:

> The history of an individual through time can be described as a succession of separate selves, which may have incompatible preferences, and may take decisions that affect subsequent selves ... Which of these selves should be granted authority over outcomes that will be experienced in the future? (Kahneman 1994: 31)

This philosophical problem is particularly salient when one makes the assumption of multiple selves. Indeed, ethical questions such as what the relationships between different selves are, or whether selves differ in the same

[1]Note that Kahneman's (2011) dual system does not necessarily commit itself to a multiple selves view, as system 1 and system 2 can be understood as different processing subroutines. The author makes it clear that system 1 and system 2 'are fictitious characters ... [they] are not systems in the standard sense of entities with interacting aspects or parts' (Kahneman 2011: 29). The same can be said for the planner-doer model of Thaler and Shefrin (1981). Although the authors explicitly make the assumption of a 'two-self economic man', they introduce their model 'by viewing the individual as an organization' (Thaler and Shefrin 1981: 394), i.e. not a literal representation of multiple selves. Yet the models of Kahneman (2011) and Thaler and Shefrin (1981) seem well compatible with a multiple selves approach to identity as (i) they assume a conflict of preferences inside the same individual and (ii) such a conflict of preference is spread over time.

[2]Those types of decision are mostly represented by the psychological phenomenon of self-control failure. Self-control failure is explained by several models of decision-making, such as quasi-hyperbolic time discounting – which encapsulates the idea that individuals have present-bias preferences (Laibson 1997) – or an axiomatic foundation of the ranking of commitment, temptation and cost of self-control (Gul and Pesendorfer 2001). In the present paper, we are however not exclusively concerned with self-control failure but with any kind of intertemporal choice that may affect one's well-being, e.g. how much to save for retirement, how to invest, whether to buy a house, whether to have children or whom to marry (Loewenstein and Thaler 1989).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

ways as individuals differ, seem to be unavoidable from a philosophical perspective. In this paper, we provide an argument for an ontological approach to personal identity when dealing with the philosophical problem of preference reversals over time. (We will refer to this problem as 'the ethical problem of identity in behavioural welfare economics'). We argue that the ontological approach to personal persistence provides a valuable alternative framework for discussing the relationship between identity and ethics in behavioural welfare economics, while such an approach has not been given particular consideration in the critical literature of behavioural welfare economics (to be discussed in sections 3, 4 and 5).[3]

The ontological approach to identity aims at determining what the nature of identity-relations over time is in order to define personal identity. It first asks what makes it the case that an individual persists over time and then aims at discussing what the ethical implications for the self would be. In other words, the aim here is to find out whether we can, first of all, actually grant identity on an ontological basis (e.g. on psychological, physical, narrative or sociological properties) and only then would we try to study what the ethical implications could be (typically whether it makes sense that the preferences of John today have more importance than the preferences of John yesterday).[4] This entails that, if such an ontological criterion (based on e.g. psychological, physical, narrative or sociological relations between John's different selves) is philosophically supported, it could inform us on which types of features (e.g. psychological, physical, narrative or sociological ones) the preferences of John today can be said to have more importance than the preferences of John yesterday. For example, assuming John's identity is defined by some psychological relations (e.g. the ability of giving sense to all of his actions by virtue of rational preferences), then an ethical rule based on John's *psychological* features (e.g. his rational preferences) would have a significant advantage over another one, e.g. one based on social norms that are external to John's psychology (to be discussed in section 6).

Within our approach, any ethical enquiry one may have about identity-relations over time is then *conditioned* by determining *ex ante* an ontological criterion of identity over time. Our general argument is the following. If we cannot first provide philosophical support for the idea that John is actually the same individual from one time to another based on some ontological criterion of identity, then it seems obscure to advance ethical rules to determine which one of the many possible preferences of John should be granted normative authority. In fact, if John is not considered as being the same individual through time, it would be as if one is evaluating the well-being of two different individuals over

[3]By behavioural welfare economics we refer to the literature which aims at extending a welfare evaluation to situations where individuals have incoherent preferences. We include in this literature Camerer *et al.* (2003), Thaler and Sunstein (2003, 2009), Bernheim and Rangel (2007, 2009) and Bernheim (2009, 2016). There is also the adversarial anti-welfarist approach of Sugden (2004, 2018) that we do not include in behavioural welfare economics because it is merely not *welfarist*. For a general review of the normative programme which aims at 'reconciling normative and behavioural economics' in response to the observation that individuals have incoherent preferences, see McQuillin and Sugden (2012).

[4]Note that the ontological approach does not impose that identity-relations are connected by moral properties – a condition that we judge to be too strict for personal persistence (to be discussed in section 4).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

time, and then the question of *individual* welfare evaluation would be at least disputable. Yet (i) if we *can* support the idea that John is the same individual from one time to another given some relations, e.g. psychological ones, then, *and only then*, (ii) does it make sense to ground normative authority on some specific features (e.g. psychological ones) that would provide ontological support for such a normative authority.

Without neglecting point (ii), our paper strongly focuses on point (i). We show that determining that individuals persist over time is actually a difficult condition to be fulfilled – a condition we need in order to move on to point (ii). In particular, we show that the literature of personal persistence (which takes the ontological approach to identity) enlightens some difficulties associated with the current alternative assumptions of the unified self proposed in the critical literature of behavioural welfare economics, which in turn have been developed in response to philosophical problems associated with the multiple-selves assumption. Additionally, we argue that the way this critical literature deals with the issues of multiple selves in behavioural welfare economics may actually lead to some complications, such as reducing the question of personal identity over time to the question of personhood ('what does it take for something to be a person?').[5]

The rest of the paper is organized as follows. Section 2 introduces an example of the ethical problem of identity applied to behavioural welfare economics (BWE). Section 3 contrasts our ontological approach to identity with related literature. Section 4 presents the framework of personal persistence by formalizing the criterion of identity over time as developed in analytic philosophy. Section 5 critically reviews the main theories of personal persistence offered in analytic philosophy. We show that most of the alternative unified-self assumptions proposed in the critical literature of BWE cope with the narrative view of personal persistence. We argue that those assumptions of the unified self are no less criticizable than the multiple-selves assumption because the narrative view is philosophically problematic as soon as we bring it to the fore of ontological questions about personal identity. After showing that there is no consensus on determining what makes it the case that individuals persist over time, our main result is that grounding any ethical rule on individual behaviour over time is actually a difficult enterprise if one does not put considerable effort in finding – first of all – a solid ontological criterion of identity. Still, assuming that such an

[5]Although distinguishing the personal persistence question ('what does it take for a person to persist from one time to another?') and the personhood question ('what is it to be a person, as opposed to a nonperson?'), we do not deny that these two questions are strictly related. As a matter of fact, one may counter that some solutions to problems posed for instance by fusion-thought experiments involve changing the understanding of persons (e.g. understanding persons as diverging time-worms). Furthermore, we do not deny that it appears meaningful, at first glance, to reduce questions about the relation between personal persistence and ethics to the question of personhood. What we want to stress here is that the personal persistence question and the personhood question concern two different (still strictly related) issues, which must not be confused. It is in fact the persistence question that animated the metaphysical discussion on personal identity over time, as it has been formulated for instance by Williams (1970) and Nozick (1981). For a presentation of several possible framings of the problem of identity (in terms of personhood, persistence and others), see Olson (2020). We come back to the relation between persistence and personhood questions in sections 4 and 5.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

ontological criterion would be consensual, section 6 suggests potential implications of the ontological criterion for the ethical problem of identity in BWE. Section 7 concludes by summarizing three advantages of our ontological approach to identity compared with the related literature.

Three useful clarifications should be stated. First, as this paper is concerned with incoherent preferences over time, we avoid focusing on *why* they can be incoherent over time (e.g. present bias), as it would add nothing relevant to the goal of the present paper. That means we merely consider incoherent preferences to be synonymous with *preference reversals*: the case in which an individual prefers A to B and then B to A at two different times. Second, we deliberately (and whenever needed) privilege the term *personal persistence* instead of identity in order to use a more accurate terminology – albeit it is not absurd to consider 'identity over time' as a synonym of 'persistence'. One reason is that identity is a potential result rather than an assumption in the literature of personal persistence. After a careful investigation of how temporal selves of individuals are related to each other, identity may, after all, not be what matters for personal persistence.[6] Third, *temporal* selves instead of multiple selves is also a privileged terminology because we are here only interested in selves which differ with respect to their *temporality*. That being said, we do not deny that the concepts of doers/hot states/system 1, on the one hand, and planners/cold states/system 2, on the other hand, may be understood as coexisting at any point of time. For example, the doer tells the individual to eat the tasty cake, while the planner prevents her at the same time from doing it. In order to be consistent with models of intertemporal choice, we consider however individual behaviour to be a matter of *preference reversal* over time, where the preference of one self overrules the preference of the other self at the point of time when the decision is being made.

2. The Ethical Problem of Identity

Our starting point is the question raised by Kahneman (1994) and Sunstein (2019) about the ethical implications of incoherent preferences: which of the many possible conflicting preferences of an individual over time should be considered as normatively relevant? Before introducing our ontological framework and arguing how it can be relevant for this problem, it is first necessary to set the ethical problem of identity in terms of multiple selves, as originally formulated. This is helpful not only to explain how the multiple-selves assumption fails to ground ethical rules for individual decision-making, but also to critically assess the 'unified-self' alternatives to such an ethical problem.

Assuming the existence of multiple selves, an individual I can be considered over time as a collection of a finite number of temporal selves $\{s_1, \ldots, s_n\}$, where s_i is a

[6]This is for example the conclusion of Parfit (1984: 215), according to which identity does not matter for persistence in terms of survival, as identity and persistence involve different kinds of relations. According to Parfit, identity is a one-to-one relation, while persistence is a one-to-many relation (for instance in terms of psychological continuity). For an introduction to Parfit's account of personal survival, see Shoemaker (2019: sec. 2.5). For critical appraisals of the implication of Parfit's theory of personal persistence for BWE, see Ferey (2011: 746–747), Hédoin (2015: 98–102) and Lecouteux (2015: 403–407).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

temporal self which exists at a given time t_i. Let $i \in \{1, \ldots, 100\}$ be the index of time. Imagine that, at t_1, s_1 preferred A to B, whereas now, at t_{100}, s_{100} changes her mind and prefers B to A.[7] As shared by Kahneman (1994) and Sunstein (2019), one central matter of concern in intertemporal choice is which one of the two (or of the many other) selves has normative authority. By 'normative authority' we mean 'moral responsibility on all the other selves'. The relationship between personal persistence and ethics is typically about the justification of one's actions (or choices) through time. John can be held responsible for some past, present and future actions only if he is actually the same individual who made, makes, or will make those actions. Also, John can justify his concerns for his future self only if this self will actually be John (and not somebody else).[8] We first consider several intuitive ethical rules and argue that they all suffer from being arbitrary. We then argue in section 3 that the ethical problem of identity constitutes a practical burden for economists for several reasons we discuss in turn. This allows us to introduce the relevance of the ontological approach to identity as an alternative approach to such a problem. We then introduce the ontological framework of personal persistence in section 4.

2.1. The present rule

One rule could state that s_{100} overrules s_1 because what matters is what happens *now*. That means that an individual is the master of her own well-being at each present self. This rule gives the present self full responsibility for her own actions. In this sense, it is well aligned with the liberal tradition of the consumer sovereignty principle in normative economics, where each s_i is considered to be the best judge of her own well-being at t_i. Instead of looking for which preferences count as normatively relevant, this rule is compatible with Sugden's (2004, 2018) approach that economists should rather promote institutional arrangements so that individuals can seek what they want — disregarding how incoherent their preferences are.

2.2. The priority rule

Another rule could state that what matters is the preference of the *first* temporal self. That means that if s_1 expresses a preference for A over B, then s_1 overrules s_{100}. This rule is likely to hold on the important condition that s_1 contracts with her (not yet existing) future selves. The contract would specify that s_1 takes full responsibility for the consequences of her preference for s_{100}, no matter what they are. Such a rule would however be endorsed by BWE only if it appears that s_1 is the far-sighted planner, cold state or system 2. If not, s_1 is making a mistake and her preference would not be considered to be normatively relevant.

The main problem with both of these rules is that they seem quite arbitrary from an external standpoint, particularly when we have no objective criterion to

[7]This example is a version of McMahan (2002: 497) formulated on our own. For a discussion of intertemporal change of preferences without multiple selves, see Ainslie (1975).

[8]For a presentation of the relationship between personal persistence and ethics, see Shoemaker (2019).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

determine what makes the normative authority of s_1 more important than that of s_{100} (and conversely). As Sunstein (2019: 79) puts it, 'there is no alternative to resorting to some kind of external standard, involving a judgment about what makes the chooser's life better, all things considered'.

2.3. The objective rule

Yet another rule would aim at shouldering this external standard by trying to determine an objective criterion that states which self or selves has/have normative authority — no matter its/their past, present and future status. There may be at least two ways to define such an objective criterion.

2.3.1. The majority criterion

One may say that if most of the selves among the finite number of all the selves prefer A to B – say $\{s_1, \ldots, s_{51}\}$ but not $\{s_{52}, \ldots, s_{100}\}$ – then 51 selves against 49 overrule whatever s_{100} states, and then the preference of A over B should be taken as the one which is normatively relevant. This rule is explicitly endorsed by Thaler and Sunstein (2003: 178), who argue that the social planner should choose a choice architecture based on the majority of individuals' expressed preferences. But if it appears that the 51 selves are the myopic doers and the 49 other selves are the far-sighted planners, economists may not consider this criterion to be reasonable.

In more specific terms, the majority selves should be expected to have a form of 'rationality', 'consciousness' or 'awareness' so that economists may reasonably think their judgement to be 'enlightened'. Thaler and Sunstein (2003: 178) recognize themselves the limits of the majority criterion, stating that the majority's choice may simply not be sufficiently informed, and that those aggregated choices may not promote the majority's well-being. Another important problem with the majority criterion is that it assumes that the selves are equally weighted, but it does not have to be so.

2.3.2. The weighting criterion

One may say that the issue is to know which of the finite number of selves is the 'supreme' one. For example, assume only s_{30} and s_{100} prefer B to A, and we discover from certain evidence that s_{30} has supreme normative authority. Using this objective rule, we could therefore account for the preference of s_{30}. This rule is endorsed by Bernheim and Rangel (2007, 2009). Although the authors do not explicitly refer to a weighting criterion, their extension of the revealed preference framework to welfare analysis requires a minimal guidance to individual well-being through choice data. On their account, only the 'fully rational' self counts for welfare analysis, even if most of the remaining selves prefer A to B. This view is however not without problems.

We may obviously question what the good reasons are to make us believe that one self should be given more weight at one time of her life instead of another. It seems that cognitive capacities are important here: some temporal selves could be eliminated from the possibility of having any normative authority – typically those

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

belonging to childhood. But then we need to determine what the 'mature and reasonable' period of one's life is. Assume that some evidence could tell us which period of one's life tends to be associated with one's informed or 'rational' choices, e.g. all temporal selves included in the set $\{s_{30}, \ldots, s_{40}\}$, and assume that we could somehow determine such an interval. It would also require that the preferences of the selves which belong to this interval remain relatively stable. But recall that the initial problem of BWE is to find a way to extend a welfare evaluation to situations where individuals have incoherent preferences. This means that if one self which belongs to the set $\{s_{30}, \ldots, s_{40}\}$ has incoherent preferences, we are no further along.

There can be other objective criteria to overcome this problem, e.g. considering the mean value of the interval (here t_{35}) as the instant when the temporal self has normative authority. The problem with any kind of objective criterion is, however, that they rule out the possibility of idiosyncratic preferences. In other terms, a general rule of individual well-being waives the possibility that individuals can perceive their 'mature and reasonable' period of life differently. For example, some may make more sense of their life as a whole at s_{67}, while others do it at s_{24}.

3. The Ontological Approach to Personal Persistence

The ethical problem of identity in BWE seems, from a practical viewpoint, quite a burden for economists to solve — particularly when they have no expertise in determining on which general criterion they can base normative authority. Perhaps only economists' personal ethical judgements can help them out, but those ethical judgements are far from being self-evident and are far from being subject to consensual agreement. For example, Bernheim (2016: 38–39) justifiably underlines the problem of 'heavily value-laden language', such as 'present bias' and 'self-control problems', which assumes that individuals have unitary preferences and equate well-being with exponential-discounted utility. Sunstein (2019: 69) also emphasizes that true happiness might be interpreted as living at the moment. That is to say, there is no a priori reason to consider the present rule as less important than any other rule. In addition, it is not impossible that ethical judgements made by economists (who are human after all), expressed in terms of preferences, are also subject to incoherence from one time to another.

A social choice alternative that we do not engage with in the present paper is to offer an ethical account of how to aggregate well-being over the temporal selves. In his discussion related to the philosophical issues of identity in BWE, Hédoin (2015: 84–88) specifically tackles the ethical problem of identity from this perspective by formulating a social welfare function of BWE. According to this function, the social planner maximizes the weighted sum of the selves' utilities of a given population with respect to an exogenous weighting parameter. The author points out the difficulty of knowing the weight of each self in the decision, especially when there are no other alternatives than making ethical judgements about which selves' decisions are considered to be more normatively relevant than others (Hédoin 2015: 89). This social choice alternative is also well recognized by Sunstein (2019), who notes that

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

> the experiencing self might have too little regard for the remembering self, but the converse is also true. It is not clear that either deserves priority. To know, we might have to make some moral judgments, or offer some account of how to aggregate well-being over time. (Sunstein 2019: 76)

Aggregating well-being over time would however ultimately yield to the arbitrary ethical rules previously discussed.[9] This is particularly the difficulty we aim to avoid by focusing on the ethical problem of identity not from a *social choice* perspective (how to aggregate individual well-being at the intrapersonal level) but, first of all, from a *personal persistence* perspective (what makes an individual *one* over time). Indeed, recall that the way we see the ethical problem of identity introduced above is that any answer coming up to determine the moral authority of an individual requires us to make an essential reference to personal persistence. To put it differently, we defend what we believe to be the reasonable view that what makes an individual morally responsible for her own action at a given time is a question that cannot be answered without first asking ourselves what makes that same individual persist over time.[10] That is to say, *the individual I can be held responsible for her past and future actions only if I is the same individual from one time to another.*

Importantly, contrary to the critical literature of BWE, which bases identity on ethical claims (to be discussed below), we add that we do not say anything on what is required to have moral responsibility. Otherwise our approach would take the same path as this critical literature. We specifically aim to avoid any ethical stance which assumes or defines the concept of a morally responsible individual in order not to bias our enquiry of what makes an individual persist over time. This 'unbiased stance' is required if we do not want to first make (arbitrary) ethical claims about what morally responsible individuals are, *and then* conclude that individuals persist from one time to another. Accordingly, we will from now on use the term 'person' to define an individual who has moral responsibility, and we shall insist that, contrary to the critical literature of BWE, we do not say anything on what is required to have moral responsibility.[11]

Our approach thus respects the following two steps. We first ask what makes an individual persist over time and only then, *if* such an ontological criterion is considered to be correct, it can provide us valuable information on which ethical rules can be supported in BWE. Indeed, on the condition that individuals persist by virtue of some relations (e.g. psychological ones), any ethical rule based on the characterization of psychological relations (for instance, the satisfaction of rational preferences) would have a significant advantage over another rule (e.g. one based on social norms). This is because such an ethical rule in BWE could be explained at an ontological level, as it would be grounded on some ontological premises. (In the case above, personal persistence consists in some given psychological relations

[9]We of course do not mean these rules to be exhaustive. Other rules that have reasons to be justified may lead to the same outcome.

[10]We say 'reasonable' because some philosophers object to the view that any ontological or metaphysical question about personal persistence is relevant to our practical moral concerns (Rovane 1998; Conee 1999).

[11]We however elaborate in the next section our position regarding the relevant question of whether personal persistence, when related to ethics, is not ultimately reduced to the question of personhood ('what does it take for something to be a person?').

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

over time). We show however in section 5 that such an ontological criterion of identity over time is far from being consensual because each view proposed in the literature of personal persistence (i.e. temporal selves either connected by psychological, physical, narrative or sociological relations) is philosophically problematic. As a consequence, the sense of our ontological approach is to show that it is difficult to pass to the second step of discussing what ethical rules of individual behaviour should be proposed when the first step (of what makes it the case that an individual persists from one time to another) has not been fulfilled *ex ante*. We however suggest some directions in section 6 if we assume the first step to be fulfilled.

In a nutshell, in order to avoid (i) intuitive reasoning that relies on common sense about which self has normative authority and (ii) a social choice approach which would consist in proposing ethical rules to aggregate intrapersonal well-being, we propose to formulate the ethical problem of identity in BWE within the framework of personal persistence from an ontological approach. This is helpful in order to advance on the philosophical enquiry of which self should be granted normative authority over another. In particular, our approach takes the definition of a criterion of identity over time seriously, as identity-relations are not required to be defined by moral properties (section 4); it underlines the philosophical problems associated with unifying the self (section 5); it suggests directions for supporting some ethical rules over other ethical rules based on ontological grounds (section 6).

4. The Criterion of Identity over Time

The problem of personal persistence consists in focusing on the criteria of identity over time.[12] A criterion of personal identity over time can be defined as the completion Φ of the following schema.

> Let x be an entity that exists at time t_i and y an entity that exists at time t_j, where $t_i \neq t_j$.
>
> Let also $Px \vee Py$, where P is the property of being a person.
>
> Then $x = y$ if and only if $\Phi(x, y)$, where $=$ is the relation of numerical identity over time, and Φ is the constitutive condition whereby the identity of x and y is determined.[13]

[12]For an introduction to personal persistence in analytic philosophy, see Buonomo (2018).

[13]This criterion of identity over time seems to violate Leibniz's indiscernibility of identicals, which states that $x = y \rightarrow \forall F(Fx \leftrightarrow Fy)$. That is, if two entities x and y are identical then they share all the same properties F. In this matter, one may find preferable to use '$x\ I\ y$' to denote the identity relation, as '$x\ I\ y \leftrightarrow x = y$' does not have to hold if we offer alternative interpretations of the I-relation (we thank one anonymous referee for pointing this out to us). Although we are very sympathetic to this remark, asking for Leibniz's indiscernibility of identicals for identity over time appears to be too strict – at least too strict for the contemporary approach of personal persistence. Indeed, one may find it reasonable to think that things persist over time (i.e. they remain numerically identical over time) even when they change their properties. For instance, consider John being 10 years old and John being 40 years old. The physical and mental properties of 10-year-old John and 40-year-old John are likely to be quite different (e.g. they have different heights, ways of thinking, etc.). But if we take Leibniz's indiscernibility of identicals to be true, 10-year-old John and 40-year-old John cannot be identical. In short, Leibniz's indiscernibility of identicals seems reasonable for *synchronic* identity, but not for *diachronic* identity.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

Numerical identity is to be distinguished from *qualitative* identity. Two things are 'qualitatively identical' if they share the same properties (e.g. two identical papers), whereas they are 'numerically identical' if they are one thing, and not more than one (e.g. the paper you are reading right now). More generally, we can say that two things are qualitatively identical if and only if they resemble each other exactly, whereas they are numerically identical if and only if they are one and the same thing. By the logical disjunction $\vee$, we mean that we do not impose the condition that both x and y remain a person from t_i to t_j. This is an important distinction to be made, as the critical literature of BWE typically assumes that (i) a person exists over time and (ii) her relationship with her future selves is necessarily related to another person. Such a conception of identity is shared by Sugden (2004), who defines a responsible agent as a human being who

> treats her past actions as her own, whether or not they were what she now desires them to have been. Similarly, she treats her future actions as her own, even if she does not yet know what they will be, and whether or not she expects them to be what she now desires them to be. (Sugden 2004: 1018)

Similarly, Hédoin (2015) defines a responsible agent as a human being which is

> responsible for all her actions and is interested in the consequences not only of her present action but also in the consequences of the future ones. (Hédoin 2015: 99)

These conceptions of identity seem to assume that there is no sense to argue about a person being an embryo in the past or a human in a vegetative state in the future since they see identity as a relationship between persons defined as e.g. rational thinkers (as in the psychological view presented below) or as a psychological unity defined by a narrative (as in the narrative view presented below). In our present framework, we however do not want to make such an essentialist assumption about persons because it would tend to reduce the question of personal persistence to (i) the question of the ontological nature of persons ('what are we really?'), or to (ii) the question of the concept of personhood ('what does it take for something to be a person?') or even to (iii) the question about the biographical identity of persons ('who am I?').[14]

We recognize that the relationship of personal persistence with these three questions may be intimately linked when ethics is involved. Some may argue that for the sake of our practical moral concerns, individuals persist over time by some psychological relation between their moral properties, which eventually constitute their identity. They may think that ethics inevitably forces us to endorse an essentialist personhood account of identity, as it seems irrelevant to be concerned with embryos or humans in a vegetative state – who by nature do not have the ability to produce any thought. But to argue that identity

[14]Some philosophers of personal persistence do impose the condition that $Px \wedge Py$ instead of $Px \vee Py$ (Swinburne 1984; Lowe 2012), which can be referred to as 'personal essentialism'.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

presupposes morality (or any kind of psychological relation) seems a very strong claim. In fact, early conditions of identity related to moral properties would make us think that a concept of the unified self necessarily has to be either psychological or narrative. (The psychological and narrative criteria of identity are to be assessed below.) We particularly think of the following identity conditions proposed by Hédoin (2015) based on Korsgaard's (1989) representation
of agency.

- *Boundary condition*. *I* can be relatively easily identified as being *I* through her agency, including intertemporal agency.
- *Narrative condition*. *I* thinks of herself as a unit of agency and can make sense of the continuity of her decisions made in the past and the decisions she is thinking to make in the future.

The narrative condition presented here cannot be an assumption of personal persistence since nothing a priori tells us what constitutes the relationship between different temporal selves. This leads us to impose the following informative conditions of a criterion of identity. A condition of identity Φ is informative if:

Non-triviality. It has a different meaning from, or at least is not logically equivalent to, the identity it constitutes.
Non-redundancy. It should be logically possible that x and y do not satisfy Φ.
Non-identity-involving. It does not presuppose the identity it should demonstrate.

Otherwise the criterion of identity is uninformative. For example, the statement '$x = y$ if and only if they are the same entity' is trivial and identity-involving because it has the same meaning and presupposes the identity it ought to demonstrate. In contrast, the narrative view, which states that '$x = y$ if and only if they can make sense of their psychological continuity' is not trivial, nor redundant, nor identity-involving.

In light of the relevance of the ontological approach, the next section critically reviews the main theories of personal persistence offered in analytic philosophy. This serves two goals. First, by highlighting the main difficulties of determining an ontological criterion of identity over time, we argue that to defend a unifying view of the self – as the critical literature of BWE does – is no less problematic than to defend the multiple-selves assumption. In particular, we show that the narrative view of personal persistence, which is the one that is mostly assumed by critical authors of BWE, faces important philosophical problems and therefore cannot be taken as an alternative to the multiple-selves assumption in BWE. Second, we argue that despite the difficulty of determining whether an individual is actually the same from one time to another – and therefore despite the difficulty of proposing ethical rules on individual behaviour that are ontologically grounded – the ontological approach can inform the ethical problem of identity in BWE. This is because the ethical rules that can be

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

proposed about intrapersonal welfare ultimately depend on *which* theory of personal persistence is considered to be correct (section 6).

5. Unifying the Self: A Complex Enquiry[15]

5.1. The psychological view

The psychological view claims that an individual is identical over time by virtue of some psychological aspects such as memories, intentions, beliefs, goals, desires and similarity of character (Parfit 1984: 204–209). This view can be stated as follows.

> Let x be an entity that exists at time t_i and y an entity that exists at time t_j, where $t_i \neq t_j$.
>
> Let also $Px \vee Py$, where P is the property of being a person.
>
> Then $x = y$ if and only if x and y are connected by some given psychological relations.

This view has had by far the most advocates, mainly because of its practical appeal: how can y be responsible for the actions of x if she is not the inheritor of x's psychology? Yet one issue concerning the psychological criterion is that it seems to imply that personhood is one's essence – i.e. that an individual cannot exist without being a person. But as previously argued, the question of personal persistence cannot be reduced to the question of personhood.[16]

Another important issue is that the notion of 'psychological' is not well specified as it may contain many aspects such as memories, intentions, beliefs, goals, desires, and important for economics, preferences. But the main concern of BWE is specifically about finding a normative approach to economics when some psychological aspects of individuals, principally individuals' preferences, are incoherent for reasons economists do not fully understand (Bernheim 2016: 13). A continuity of incoherent preferences would then need to justify how these incoherent preferences are actually continuous, which seems a challenge one

[15]The present section is largely based on the taxonomy of Shoemaker (2019), who reviews the main theories of personal persistence and discusses their various ethical implications. Whenever needed, we associate each view of the unified self proposed in the critical literature of BWE with the underlying theory of personal persistence it endorses.

[16]One may object that this is not necessarily true for all psychological accounts of persistence, as it is the case for the Parfitian reductionist account. Indeed, one may argue that Parfit (1984) endorses a reductionist view of personhood, which should not be confused with the assumption that personal identity is primitive. However, using this Parfitian counterexample as a general defence for psychological accounts of personal identity is actually misleading. This is because the Parfitian account is a very specific and non-standard account of personal persistence, which is characterized by the (very non-standard) rejection of the assimilation between 'personal identity' and 'personal persistence' – commonly summarized in Parfit's famous sentence 'identity does not matter for survival'. Given this account, the ethical claim of personal identity is prioritized, whereas the ontological question follows after that. Our intention here is not to discuss Parfit's approach, but rather to reject the use of Parfit's reductionist psychological approach as the standard approach for psychological views on personal identity. We stress that Parfit provides a very specific and not generalizable argument to face the first objection against psychological views. On the revisionary aspects of Parfit's theory, see Martin (1998: 15) and Rovane (1998: 11).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

cannot face without relying on some essentialist assumptions. These essentialist assumptions would be e.g. the existence of a far-sighted planner, or the existence of true preferences, which is an essential property of what some authors refer to as the 'inner rational agent' (Infante *et al.* 2016). But can we assume that one's identity is located in one's psychological property that neither behavioural economists nor neuroeconomists are able to locate?

Furthermore, it seems presumptuous to argue not only that one is constituted by an inner rational agent which is the source of one's normative authority (and potentially also one's identity), but also that one cannot make way for other 'psychological roles' when one makes a decision, for example making a decision as a parent or a wife.[17] It is thus not surprising that no one has so far proposed a convincing account of the psychology of the inner rational agent. This rational agent supposedly has true (or latent) preferences that are accessible under conditions where she is undistorted from cognitive biases (Sugden 2015; Lecouteux 2016). But it remains a mystery whether those true preferences are actually produced or assumed to exist exogenously – as in the neoclassical consumer choice theory.

Arguably, the psychological view may miss something that the critical authors of BWE have argued to be important for identity: the *meaning* one attributes to one's own psychological relations (e.g. desires, intentions, life goals). In his reconstruction of normative economics without the concept of preference, Sugden (2018) argues that it is only required to assume that an individual is a 'responsible agent', who can give a continuous meaning to each of her own actions at any given period of her life. This seems to avoid the practical burden of justifying the circumstances under which the selves have normative authority – that is, the circumstances under which they do not make mistakes.[18]

Sugden's (2004) view is also similar to the way Hédoin (2015) and Dold and Schubert (2018) interpret identity in normative economics. We discuss their narrative view of identity below. Before doing so, let us briefly introduce another approach that is compatible with our position that personal persistence cannot be reduced to personhood, but which, at the same time, claims that identity is not a matter of a psychological relation.

[17]We suspect some readers to answer that point by saying that a decision of a parent or a wife is outside the scope of economic theory, and that it is therefore pointless to talk about a behaviour which is not even taken care of by the theory. We believe such an answer to be misleading if we follow leading behavioural economists who take any sort of behaviour to be explained by intertemporal choice, such as how much schooling to obtain, whom to marry, or whether to have children (Loewenstein and Thaler 1989: 181). Thaler and Sunstein (2009) consider any kind of life situation as examples to justify libertarian paternalism, such as avoiding the temptation of eating too much of the cashew-nuts bowl before dinner (Thaler and Sunstein 2009: 40). Camerer *et al.* (2003: 1244–1245) even consider the decision of committing suicide as a case for policymaking in their proposition of asymmetric paternalism. All these examples involve intertemporal choices that go beyond the archaic delimitation of economics to a limited set of decisions such as consumption, production, saving and investment. Although we do not particularly support the rhetoric of libertarian paternalism of justifying nudging in any kind of life decision (such as not eating cashew nuts before dinner), we seriously follow the view of behavioural economists who think that economic theory can explain any kind of choice that involves intertemporality.

[18]For theoretical frameworks of BWE which aim at identifying mistakes, see Köszegi and Rabin (2007), Beshears *et al.* (2008) and Bernheim (2016).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

5.2. The physical view

Philosophers who are unsatisfied with the psychological view argue that it should not be a matter of fact that personhood is the essence of an individual, simply because it is hard to deny that an embryo who becomes an individual and then a human in a vegetative state is not the same individual (Olson 1997; Hershenov 2005). Instead, it would perhaps be more convincing to define a continuous individual with respect to her physical properties. The physical view can be stated as follows.

> Let x be an entity that exists at time t_i and y an entity that exists at time t_j, where $t_i \neq t_j$.
>
> Let also $Px \vee Py$, where P is the property of being a person.
>
> Then $x = y$ if and only if x and y are connected by some given physical relations.

Physical relations are not necessarily located in the brain. More generally, the physical view states that physical continuity, which constitutes the biological organism of a human being, is the constitutive condition for personal identity over time, and then for her persistence.

The physical view seems nonetheless far less appealing from an ethical viewpoint because it seems irrelevant to locate identity in a physical property that has per se no function of reasoning or consciousness. But advocates of the physical view typically argue for a biological continuity between all the stages of the body as a whole, e.g. from an embryo to a rotten skeleton. For a person to be held morally responsible, its biological relationship should then have the function of producing thoughts that can be assimilated to a moral continuity. Yet neither an embryo nor a rotten skeleton is able to produce any thought.

But this is not the main issue. Assume, for the sake of argument, that we are here only concerned with a perfectly healthy middle-aged person. Assume her cerebrum is transplanted into a different living body, and that the resulting person is psychologically exactly the same as the first person (Olson 1997: 43–51; DeGrazia 2005: 51–54). By virtue of biological continuity, advocates of the physical view would argue that the cerebrum-less donor remains the same person, and that the other cerebrum-receiver is an imposter. But as Shoemaker (2019: sec. 2.2) argues, this seems hard to believe.

There are, of course, some replies to this thought experiment (Olson 1997: 70; DeGrazia 2005: 60–61) that are pointless to discuss here. What is important to emphasize is that the physical view seems unappealing to the ethical problem of identity, and that it may provide a practical argument for tenants of personhood essentialism. In any case, since the physical view is not endorsed by any view we are aware of in economics (except perhaps by some neuroeconomists), we will spend no more time discussing it.[19]

[19]In response to Lecouteux's (2015) criticism, according to which libertarian paternalism presents an implausible model of identity, Sunstein (2015: 527) mentions the possibility of considering the physical view as an alternative to Parfit's reductionist account of identity. In his words, 'consider a competing

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

5.3. The narrative view

The psychological condition of identity seems fundamental to ethics. Indeed, the ethical problem of identity introduced in section 2 seems to ask the following question: 'what psychological characteristics are attributable to the overall individual?' It is then not surprising that the narrative view has the most advocates in the critical literature of BWE. This view can be expressed as follows.

> Let x be an entity that exists at time t_i and y an entity that exists at time t_j, where $t_i \neq t_j$.
>
> Let also $Px \vee Py$, where P is the property of being a person.
>
> Then $x = y$ if and only if x and y are connected by some self-told narrative relations.

In other words, 'x and y can make sense of their psychological continuity'. The narrative view departs from the psychological view in the sense that it gives a *meaning* to the psychological relations of e.g. memories, desires and preferences. In comparison with the psychological view, it does not take the memories, desires and preferences of one's life as merely isolated events, but it weaves them together and gives them some form of coherence and intelligibility that they would not otherwise have. We can thus see identity as a story of one's life according to the circumstances of one's life (Schechtman 1996: 96–99).[20]

According to Schechtman (1996), what is more appropriate for the relation between identity and ethics is not the condition of *numerical* identity, as we formulate it in our framework by the relation $=$, but the condition of *characterization* of one's identity. That is, the question would not be 'what are the conditions under which an individual remains one through time?' but rather 'what are the conditions under which various psychological characteristics, experiences, and actions are properly attributable to a person?' In other words, the question would be 'what makes the past or future states a person is specially concerned about *hers*?' We are then back to the essentialist assumption of personhood.

Like Schechtman, the concept of identity endorsed by Sugden (2004, 2018), Hédoin (2015) and Dold and Schubert (2018) seems to prioritize the *characterization* condition before the *numerical* condition. These views may presuppose the numerical condition, but do not give an account for it. In the narrative view, what makes a psychological characteristic attributable to a person

view: In virtue of the *relevant* physical facts (for example, the same body, most importantly including the same brain), Oscar remains the same person over time' (his emphasis). We do not however believe that this point is raised seriously by Sunstein (at least not by virtue of avenues of future research on justifying libertarian paternalism), considering that his question of whose self should be attributed normative authority in Sunstein (2019) would otherwise be self-defeating. To be specific, if any unified view of the self is a priori endorsed (e.g. physicalism), there is no point to assume the multiplicity of the selves with respect to their temporality. For a defence of the physical view – often known as *animalism* – see Noonan (1998), Olson (2003) and Blatti and Snowdon (2016).

[20]Note that the capacity of an individual to provide a self-told narrative can be rooted in things other than psychological abilities, such as culture and social practices. In this sense, the narrative view is not only closely related to the psychological view but also to the sociological view (to be discussed below).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

(and thus a proper part of her true self) is its 'correct' incorporation into the self-told story of her life (MacIntyre 1984, 1989; Taylor 1989; Schechtman 1996, 2020; DeGrazia 2005). Although it could appear that x and y are numerically different, the idea is that they can still be unified by – what we intend to call – a phenomenological feature of their self-told narrative. While this theory of identity is appealing from the viewpoint of ethics, it has however some serious flaws that we consider in turn.

First, it is left unclear why we need to tell ourselves a certain story in order to attribute to ourselves a unity of the events in our life, taken as a whole. As Shoemaker (2019: sec. 2.3) puts it, we may have a robust psychological unity without having told ourselves any kind of story – and this story we are telling ourselves might simply be wrong (or in accordance with the vocabulary of behavioural economics, 'biased'). We might also want this narrative to be seen from a third-person standpoint, i.e. independently from the first-person standpoint. But the continuous self can constantly revise her own self-story. Another point raised by Shoemaker (2019) is that narrative unity seems to be a fuzzy condition of identity because it is left unclear that 'intelligible' actions (or choices) are those for which the individual is morally responsible. As he argues, 'actions of children and the insane can be perfectly intelligible – even intelligible within some kind of narrative structure – without being those for which the agents are accountable' (Shoemaker 2019: sec. 7). In the ethical problem of identity introduced in section 2, many would find it unreasonable to attribute normative authority to the childhood selves, although the narrative of one's childhood may actually have the strongest structure among all one's other narratives. That is, we would intend to think that it is not the interval of the temporal selves during childhood which is normatively relevant, but everything that happens afterwards. But tenants of narrativity would argue that we should account for *all* selves of one's life, and then weave their preferences together by some overall narrative. We however suspect many economists to reject this view because a form of 'reason' or 'rationality' seems far more appealing to characterize moral accountability than a narrative one.

Second, the narrative view endorsed in the critical literature of BWE leads to the following disturbing paradox. Authors who reject the assumption of multiple selves also reject the idea that a far-sighted planner exists by virtue of her rational capacities to know what is best for her. But at the same time, they account for a narrative unity which supposes that one can – through some psychological process that is, by the way, also left unexplained – make an 'intelligible' (not to say 'rational') story by which all one's choices are collected into a unified narrative. It is true that the continuous individual as presented in the narrative view does not presuppose that one has coherent preferences at each period of time. As Sugden (2004, 2018) puts it, the individual can have incoherent preferences and yet – we add, based on a mysterious psychological ability – make 'sense' of this continuity. This would however assume that there exists a supreme self (as the one in the weighting criterion of section 2) that can indeed make sense and collect those incoherent preferences into a coherent (or intelligible) story. But this cannot be so, because the narrative view states that *all mental states of one's life, once gathered together meaningfully, make it the case*

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

that the self is unified. Who this supreme 'phenomenological' self is remains nonetheless an open question. In our view, it is merely a soul or a ghost. The characterization condition of Schechtman (1996) thus becomes unappealing to us because the unity of a narrative – as we have just argued – requires a unity of the self who tells such a story. This ultimately presupposes strict *numerical* personal identity (MacIntyre 1984: 206–208; DeGrazia 2005: 114). The point is that in the narrative view, one cannot be a person who has an identity unless one weaves the various experiences of one's life together into a unified story. But as Shoemaker (2019: sec. 2.3) puts it, 'the identity of that subject of experiences must be preserved across time for its experiences to be so gathered up'. This explains our commitment to the numerical identity condition as previously presented.

This also explains why we consider the condition of $Px \vee Py$ instead of $Px \wedge Py$. The explanation goes as follows. Assume $Px \wedge Py$, and then that the identity question is reduced to the question of personhood. This would mean that individuals persist only by virtue of being persons. A person, broadly defined, is an individual who has the ability of being morally responsible. Identity is thus reduced to an individual who has moral thoughts, and the question of personhood would then require an answer regarding what makes it the case that an individual is a person. This account of identity would necessarily cope with the psychological view of identity, according to which '$x = y$ if and only if x and y are connected by some given psychological relations'. By providing for continuity in those psychological relations, the narrative view unifies the many experiences of one's life. But it also requires that this same individual, who can give meaning to such a psychological continuity, persists through time (such as e.g. an immaterial soul or a ghost), apart from the living entity at each t_i who may have incoherent preferences. Consequently, the narrative view would be formulated as '$x = y$ if and only if x and y are the same unified person who give psychological meaning to the actions of x and y'.

But 'x and y being the same person' violates our informative criterion, according to which a criterion of identity cannot be trivial nor presuppose the identity it should demonstrate (see section 4). It follows that the narrative condition of persistence would not appear as a strong candidate for an ontological criterion of identity, since it presupposes the identity it is supposed to explain.[21]

In his theory of the individual in economics, Davis (2011) provides what we judge to be a more compelling framework for the criterion of identity because he keeps the numerical condition. The author formulates the following two criteria of identity.

- *Individuation*. Individuals can somehow be successfully represented as distinct and independent beings.
- *Reidentification*. Individuals that have already been shown to be distinct and independent in some conception of them can be reidentified as distinct and independent in those same terms across some process of change.

[21]We add that there are more philosophical issues related to the narrative view that we are constrained not to discuss here. For a recent assessment of the narrative view, see Olson and Witt (2019).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

As Gallois and Hédoin (2017) put it, the boundary and narrative conditions of Hédoin (2015) can be seen as respective answers to the individuation and reidentification criteria of Davis (2011) – although (as we have previously stated) we provide awareness that narrativity as an ontological criterion of identity is at least problematic, since it presupposes the identity it is supposed to explain. In our view, the reidentification criterion is a more acceptable criterion of identity since it does not presuppose the narrative nor the personhood condition. In comparison to our framework, $=$ can be understood as our individuation criterion (the fact that x and y are numerically the same at different moments of time) and Φ as our reidentification criterion (the condition that makes x and y being numerically the same individual at different moments of time).

5.4. The sociological view

The sociological view (Schechtman 2014) can potentially conciliate two problems of the physical view, on the one hand, and of the psychological and narrative views, on the other hand.[22] Recall that the physical view goes too far into essentialism, and that the psychological and narrative views oppositely deny the constitution of one's identity that goes beyond one's psychology. What is nonetheless common to the biological, psychological and narrative views is that they represent identity from a *first-person* standpoint. But for each of these views, neither the social status of identity – how individuals are contextualized in their social environment – nor the story of their life told from a *third-person* standpoint is suggested. The sociological view can instead be formulated as follows.

Let x be an entity that exists at time t_i and y an entity that exists at time t_j, where $t_i \neq t_j$.

Let also $Px \vee Py$, where P is the property of being a person.

Then $x = y$ if and only if x and y are connected by some sociological relations.

According to this view, social relationships produce persistence. A reductionist sociological view can be based, for instance, on the idea that assigning social identification numbers is what grounds personal persistence, and that is what constitutes persons over time. If social systems constitute personhood, this view can be informative on how individuals' normative authority can be determined (to be discussed in section 6).

According to Schechtman (2014: Ch. 5), human beings are characterized not only by virtue of their biological and psychological features, but also by virtue of their socially shaped capacities. Schechtman considers that human beings evolve in their contextual environment – a family, a community, a nation – where these social features are essential properties of what characterizes an individual. That is to say, every social factor that characterizes a human being born in a given

[22]Schechtman (2014) calls it the 'person-life view' and Shoemaker (2019) the 'anthropological view'. As we believe 'sociological' to be a term that better contrasts with the previous three views of identity, we prefer the latter term over the two former.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

environment (her culture, norms, habits) forms her ontological unit that gradually becomes responsible for and concerned with its own future (Shoemaker 2019). Such a responsible unit is no different from the embryo from which she evolved, and her unity as a being remains after she dies since funerary customs preserve the identity of buried rotten skeletons. Schechtman's view of identity is similar to that of Davis (2011). Davis (2011: Ch. 3) provides an extensive account of 'socially embedded individuals' but in contrast with Schechtman, he precisely accounts for both the narrative and sociological views of personal persistence:

> [individuals'] self-narratives about how they themselves look upon their choices trade in the language and meanings of this social discourse and cannot be understood apart from it … From this perspective, self-narratives are both highly individualized and highly institutionalized accounts people produce to track how they see their own capability development pathways. (Davis 2011: 213)

Davis particularly criticizes the model of social identity of Horst *et al.* (2007) for not considering individuals' preferences to be endogenously determined by their social background. He argues that those preferences have no reason to be exogenous because individuals' preferences are constantly changed by an 'individual-to-society' relationship he characterizes by the notion of capability (Nussbaum and Sen 1993). We thus interpret the concept of identity of Davis as a *hybrid* between the narrative and sociological views.[23]

Another eminent account of the sociological view is given by Ross (2005, 2014). Ross (2005: Ch. 8) develops a 'narrative-sociological' approach, where individuals progressively build their characters through strategic interactions, relying on institutions (especially language). In his investigation about what normative economics is fundamentally about, Ross (2014) defends a strong connection between the two disciplines of economics and sociology, claiming that 'individuals … are products of social structure, not components into which social structure can be analyzed' (Ross 2014: 286). According to him, the fundamental ontology of economics is not individuals but *markets*. He argues that 'the principles of normative decision theory … [are] more closely approximated by … groups of people making choices in particular kinds of institutional contexts' (Ross 2014: 36) than by individuals making choices in relative isolation.[24]

We think the sociological view may be an interesting candidate for the ontological criterion of identity we have so far discussed, especially when a consequent body of empirical studies supports the view that individual

[23]Some may argue that the relation between narrative and sociological views is even closer, so close that these accounts cannot be dissociated. For instance, one may argue that every narrative view should be sociological at the same time, as it is difficult to see how someone could build her own narrative in a purely introspective manner. Although this argument would deserve a longer discussion, let us accept it for the sake of argument. Even in this case, it would not follow that every sociological account is narrative, and then it would not follow that these views cannot be dissociated.

[24]Such a 'narrative-sociological' account of identity can also be found in Sugden (2018), who sees markets as institutional arrangements.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

preferences are socially shaped (Chen and Li 2009; Benjamin *et al.* 2010). Insofar as the ethical identity problem of BWE is considered, note that it is implicitly formulated from a third-person standpoint. In Sunstein's (2019) ethical concern of libertarian paternalism, the question is 'which of the several selves has/have normative authority from the *social planner's* standpoint?' (assuming the social planner is the ultimate judge of one's well-being). The social planner is however always represented as another single individual (or at best, a group of individuals), but not as the society taken as a whole.

We suspect that BWE does not implicitly assume the sociological view of identity because it would introduce sensitive debates about whether individuals should conform to norms. This is paradoxically already proposed in Thaler and Sunstein (2009), who consider the habits of saving more and eating healthy as morally desirable.[25] Furthermore, if norms were already fully embedded in economic behaviour (e.g. it is a western norm to eat healthily, to exercise and not to smoke), then the social planner would have no role in accounting for individuals' preferences which deviate from 'good behaviour', e.g. for self-control failures.

6. On the Ethical Implications of the Criterion of Identity over Time for Behavioural Welfare Economics

In the previous section we emphasized that most of the unified views discussed in the literature of personal persistence (especially the *narrative* view) are philosophically problematic, and that the identity criterion is better defined by a numerical instead of a characterization condition. The first conclusion is that the unified-self assumption (particularly defended by the narrative condition in the critical literature of BWE) appears no less problematic than the multiple-selves assumption. What is even more important for our ontological account of personal persistence is that if one does not give stronger arguments for the narrative view of personal persistence, any ethical rule based on one's narrativity would be considered as problematic in the ethical problem of identity in BWE.

As we have previously argued, recall that ontology is important because if we cannot first say that one remains the same from one time to another based on some ontological criterion of identity over time, then it seems pointless to attribute ethical rules to one's actions through time. On the contrary, for an individual to be held morally responsible for her own actions over time, one necessary condition seems to be that she actually remains the same from one time to another. Accordingly, the literature of personal persistence is specifically devoted to provide an answer to the ontological question of what makes it the

[25]Perhaps the most sophisticated attempt to address multiples selves in BWE is Bénabou and Tirole (2002, 2003). The authors represent individuals as a collection of multiples selves located at any point in time, who are '*imperfect Bayesians*' (Bénabou and Tirole 2002: 898) (their emphasis). In their representation of individual identity, they account for the psychological phenomena of self-confidence and personal rules but acknowledge that such a psychological account cannot be complete without giving attention to 'interacting with others' and to the 'social environment' (Bénabou and Tirole 2003: 159). For a critical review of Bénabou and Tirole (2002, 2003) as an incomplete account of the sociological view, see Davis (2011: Ch. 3).

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

case that an individual persists through time, but as we have shown, no ontological criterion over time creates consensus among philosophers.

Another important aspect that has been left out so far is the potential implications of a given criterion of identity over time for welfare analysis. In other words, *assuming* that an ontological criterion of identity would be considered as being correct/consensual, what would be the implications for the ethical problem of identity in BWE? The aim of this section is to focus on this last point. In fact, the specific ontological criterion over time we consider to be correct may lead us to different normative recommendations. Since we have reviewed in the previous section four ontological criteria of identity over time (the psychological, the physical, the narrative and the sociological criteria), we briefly discuss some of their possible ethical implications in turn.

Assume the psychological criterion is correct, that is, that x and y are the same person over time because they are connected by some psychological relations such as memories, beliefs, preferences or desires. Then, given the view that individuals persist by virtue of the fact that they are connected by (say) preferences that satisfy some properties, it seems that any ethical rule based on the characterization of 'true' preferences (such as in BWE) would have a significant theoretical advantage over other rules. This is because they could be explained by referring to something on an ontological level, i.e. the fact that personal persistence is a matter of preference continuity. To be clear, we do not want to say that the psychological view *justifies* BWE, but that on the condition that individuals persist because of some kind of psychological structure, then and only then does it make sense to build ethical rules based on their preferences.

More difficult would be to think about what the ethical implications of the physical view would be for BWE (in particular) and normative economics (in general). This is because normative economics is typically about evaluating states of affairs or recommending public policies that are either based on subjective criteria located in individuals' mind (typically preferences), or on objective criteria that are external to their body (such as enhancing levels of security, employment, freedom, etc.). As we have argued in the relevant subsection, since the physical view is not endorsed by any view we are aware of in normative economics – but it could potentially be of relevance to neuroeconomists, who might see connections between identity and the physical body through neural activity – we leave the ethical implications of this view aside.

Assume now the narrative view is correct, that is, that x and y are the same person over time because they are connected by some self-told narrative relations. Then, given the view that individuals persist because they are connected by some narrative structure in which they can make sense of their actions through time, it seems that any ethical rule based on such a narrative structure makes sense, because it refers to an ethical rule based on something that is ontologically grounded: the ability an individual has to weave memories, desires and preferences together and give them some form of coherence and intelligibility that they would not otherwise have. As Sugden's (2004) approach is based on the assumption that the individual is a responsible person over time, and as his approach seems to cope with the narrative view, we do not want to say that the

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

narrative view necessarily *justifies* Sugden's (2004) approach. What we want to say is that the correctness of the narrative view would represent a solid ground for building ethical rules based on his approach.

Assume now the sociological view is correct, according to which x and y are the same person over time because they are connected by some sociological relations. Granting ethical rules based on sociological features such as norms and habits is common to our way of living. In this case, given the view that individuals persist by virtue of the fact that they are connected by norms and habits, it seems that any ethical rule based on the characterization of some institutions (such as those that promote the free market) would have a significant theoretical advantage over other rules. This is because such ethical rules could be explained at an ontological level, i.e. in this case, that personal persistence is a matter of sociological features such as norms and habits. Again, we do not aim at justifying a normative approach based on a defence of a sociological account of identity. Rather, we intend to show that endorsing a sociological account of personal identity would provide an explanatory advantage to any account of ethical rules based on sociological features (such as capabilities, as in Davis' account on identity).

To make sense of our ontological approach for BWE, imagine the example of John, a young man who believes he is living an existential crisis. Imagine that John is not sure about what he wants, and therefore experiences strong preference reversals over time. In particular, he is unsure about the idea of continuing his academic career (A), or of dropping everything and travelling around the world (B).

If the psychological criterion is accepted and we observe that he has a preference for A over B at t_1 and B over A at t_2, then maybe by virtue of 'rational preferences' – as argued by proponents of libertarian paternalism and BWE – we could say that he has made a mistake at t_2 by quitting his job at the university, and that he figured out that his trip around the world is in fact not what he truly wants. In this case, the psychological criterion of identity could tell us that considering he is the same individual from t_1 to t_2 with respect to some psychological relations (e.g. rational preferences), then and only then could it make sense to say that he prefers A to B at t_1 and t_2. If the narrative condition is accepted, we can make sense of John's story: from when he wanted to have an academic career at t_1 to when he wanted to quit everything to live a more adventurous life at t_2. In this sense, it could make sense that in his storytelling, if he prefers A to B at t_1 and B to A at t_2, then A is better for him at t_1 and B is better for him at t_2. Assume now the sociological view is correct, and we observe that he has a preference for A over B at t_1 and B over A at t_2. If we can say that he feels social pressure at his age for not continuing his academic career and that his personal values lead him to leave everything for a trip around the world, then because he can be defined as the same person through time by virtue of some sociological properties – say, the habits of his cultural environment – then and only then does it make sense to argue that B over A is best for him at t_1 and t_2, assuming for example that his personal values should here prevail over his cultural environment.

Note that we do not explain which preference is best for John – either A over B or B over A – but that the ontological criterion of identity is *informative* (if not a

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

necessary condition) towards what could be said about which preference of *A* over *B* or *B* over *A* makes John better off over time.[26]

7. Conclusion

In this paper we proposed an alternative approach to the ethical problem of identity in BWE, which consists in considering ontological questions of personal identity as fundamental in order to advance on such a philosophical problem. In order to contrast our approach with related literature in economics-and-philosophy, we called our proposition the 'ontological approach' to personal identity in BWE. According to this approach, ethical questions on personal identity in BWE can be informed on the basis of the answer given by ontological questions about personal persistence – in our case: what does it take for an individual to persist from one time to another? To achieve this aim, we introduced the way personal persistence is framed in analytic philosophy and then presented the main theories of personal persistence in the current analytic debate. This was to show two important results. First, since unifying the self is a complex enquiry (especially in the narrative view), any ethical stance about how to evaluate individual welfare over time can hardly be defended on the basis that individuals are connected over time by some given properties (e.g. narrative ones). As a result, the unified-self assumption appears no less problematic than the multiple-selves assumption in BWE. Second, on the condition that an ontological criterion is judged to be correct, it can potentially inform us on which normative recommendations can be undertaken in treating the ethical problem of identity in BWE. For example, if the psychological view is correct, it would provide a significant support for defending a psychological ethical rule of individual behaviour over time (compared with another ethical rule, which is based for example on social norms).

We see three main advantages of our approach. First, it offers the opportunity of dealing with personal identity in BWE without requiring any *ex ante* commitment to a specific normative account of personal identity. Instead, it can provide support for a normative recommendation, *which follows* the ontological research on the way individuals persist through time. In this way, we provide a novel approach to the related literature, the latter being rather concerned by an *ex ante* commitment to a specific normative account of personal identity (typically moral responsibility defined in terms of narrativity). Second, it prioritizes the enquiry of taking on the philosophical problems of identity over the question of individual welfare evaluation. As we have stated, if we cannot confidently maintain that John is the same individual from one time to another (or in other words, if the unified-self assumption is philosophically problematic), then we should perhaps first allocate

[26]There is of course Hume's classic 'is-ought' problem, according to which one cannot derive ethical judgements from ontological principles. For example, even if we consensually agree that the sociological view is the best account we can find of personal persistence, any ethical claim derived from this personal persistence view is another philosophical question to be solved. Although we are well aware of this potential objection, an assessment of Hume's 'is-ought' problem leads us to another vast literature that is outside the scope of the present paper.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

our efforts in finding a better account of the unified self (e.g. the sociological view) *before* even discussing which ethical rules of identity over time should prevail over others. Third, an ontological approach to identity can both offer a solid ground for proposing ethical rules based on individual behaviour and lead us to different ethical implications. If it appears that John can be unified through time by some given relations (either psychological, physical, narrative or sociological ones), then it can make sense to attribute a meaning to all of his actions over time based on those given relations. But if 10-year-old John and 40-year-old John appear to be (ontologically) different, then it is left unclear under which ethical rule should John's preferences over time be granted normative authority. In these aspects, our ontological approach can support and eventually steer some discussions within the identity debate in BWE.

Acknowledgements. We thank three anonymous referees and corresponding editor Itai Sher for their valuable comments. We also thank John Davis, Cyril Hédoin and Marya Schechtman for useful feedback on early drafts of this paper, as well as the audience of the 15th International Network for Economic Method conference in November 2021 for their remarks. Finally, we are grateful to Pierre Van Zyl for proofreading. All mistakes remain ours.

Competing Interests. None.

References

Ainslie G. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological Bulletin* **82**, 463–496.

Bénabou R. and J. Tirole 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics* **117**, 871–915.

Bénabou R. and J. Tirole 2003. Self-knowledge and self-regulation: an economic approach. In *The Psychology of Economic Decisions. Volume One: Rationality and Well-Being*, ed. I. Brocas and J. Carrillo, 137–167. Oxford: Oxford University Press.

Benjamin D.J., J.J. Choi and A.J. Strickland 2010. Social identity and preferences. *American Economic Review* **100**, 1913–1928.

Bernheim B.D. 2009. Behavioral welfare economics. *Journal of the European Economic Association* **7**, 267–319.

Bernheim B.D. 2016. The good, the bad, and the ugly: a unified approach to behavioral welfare economics. *Journal of Benefit-Cost Analysis* **7**, 12–68.

Bernheim B.D. and A. Rangel 2007. Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review* **97**, 464–470.

Bernheim B.D. and A. Rangel 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics* **124**, 51–104.

Beshears J., J.J. Choi, D. Laibson and B.C. Madrian 2008. How are preferences revealed? *Journal of Public Economics* **92**, 1787–1794.

Blatti S. and P.F. Snowdon, eds. 2016. *Essays on Animalism: Persons, Animals, and Identity*. New York, NY: Oxford University Press.

Buonomo V. 2018. A brief guide to personal persistence. In *The Persistence of Persons: Studies in the Metaphysics of Personal Identity over Time*, ed. V. Buonomo, 7–18. Neunkirchen-Seelscheid: Editiones Scholasticae.

Camerer C., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin 2003. Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'. *University of Pennsylvania Law Review* **151**, 1211–1254.

Chen Y. and S.X. Li 2009. Group identity and social preferences. *American Economic Review* **99**, 431–57.

Conee E. 1999. Metaphysics and the morality of abortion. *Mind* **108**, 619–646.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

Davis J.B. 2011. *Individuals and Identity in Economics*. Cambridge: Cambridge University Press.
DeGrazia D. 2005. *Human Identity and Bioethics*. Cambridge: Cambridge University Press.
Dold M.F. and C. Schubert 2018. Toward a behavioral foundation of normative economics. *Review of Behavioral Economics* **5**, 221–241.
Ferey S. 2011. Paternalisme libéral et pluralité du moi (Libertarian paternalism and multiple-selves). *Revue Économique* **62**, 737–750.
Gallois F. and C. Hédoin 2017. From identity to agency in positive and normative economics. *Forum for Social Economics* **50**, 1–17.
Gul F. and W. Pesendorfer 2001. Temptation and self-control. *Econometrica* **69**, 1403–1435.
Hédoin C. 2015. From utilitarianism to paternalism: when behavioral economics meets moral philosophy. *Revue de Philosophie Économique* **16**, 73–106.
Hershenov D. 2005. Do dead bodies pose a problem for biological approaches to personal identity? *Mind* **114**, 31–59.
Horst U., A. Kirman and M. Teschl 2007. *Changing identity: the emergence of social groups*. Princeton, NJ: Institute for Advanced Study, School of Social Science, Economics Working Papers.
Infante G., G. Lecouteux and R. Sugden 2016. Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* **23**, 1–25.
Kahneman D. 1994. New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics* **150**, 18–36.
Kahneman D. 2011. *Thinking, Fast and Slow*. London: Penguin Books.
Korsgaard C.M. 1989. Personal identity and the unity of agency: a Kantian response to Parfit. *Philosophy & Public Affairs* **18**, 101–132.
Köszegi B. and M. Rabin 2007. Mistakes in choice-based welfare analysis. *American Economic Review* **97**, 477–481.
Laibson D. 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* **112**, 443–477.
Lecouteux G. 2015. In search of lost nudges. *Review of Philosophy and Psychology* **6**, 397–408.
Lecouteux G. 2016. From homo economicus to homo psychologicus: the Paretian foundations of behavioural paternalism. *Œconomia* **6**, 175–200.
Loewenstein G. and R. Thaler 1989. Anomalies: intertemporal choice. *Journal of Economic Perspectives* **3**, 181–193.
Lowe E.J. 2012. The probable simplicity of personal identity. In *Personal Identity: Complex or Simple?*, ed. G. Gasser and M. Stefan, 137–155. Cambridge: Cambridge University Press.
MacIntyre A. 1984. *After Virtue*. Notre Dame, IN: University of Notre Dame Press.
MacIntyre A. 1989. The virtues, the unity of a human life and the concept of a tradition. In *Why Narrative?*, ed. S. Hauerwas and L. G. Jones, 89–110. Grand Rapids, MI: W.B. Eerdmans.
Martin R. 1998. *Self-Concern: An Experiential Approach to What Matters in Survival*. Cambridge: Cambridge University Press.
McMahan J. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
McQuillin B. and R. Sugden 2012. Reconciling normative and behavioural economics: the problems to be solved. *Social Choice and Welfare* **38**, 553–567.
Noonan H.W. 1998. Animalism versus Lockeanism: a current controversy. *Philosophical Quarterly* **48**, 302–318.
Nozick R. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
Nussbaum M.C. and A. Sen, eds. 1993. *The Quality of Life*. New York, NY: Oxford University Press.
Olson E.T. 1997. *The Human Animal: Personal Identity without Psychology*. Oxford: Oxford University Press.
Olson E.T. 2003. An argument for animalism. In *Personal Identity*, ed. R. Martin and J. Barresi, 318–334. Malden, MA: Blackwell.
Olson E.T. 2020. Personal identity. In *Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), ed. E.N. Zalta. <https://plato.stanford.edu/archives/win2020/entries/identity-personal/>.
Olson E.T. and K. Witt 2019. Narrative and persistence. *Canadian Journal of Philosophy* **49**, 419–434.
Parfit D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
Ross D. 2005. *Economic Theory and Cognitive Science: Microexplanation*. Cambridge, MA: MIT Press.
Ross D. 2014. *Philosophy of Economics*. London: Palgrave Macmillan.

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press

Rovane C. 1998. *The Bounds of Agency*. Princeton, NJ: Princeton University Press.
Schechtman M. 1996. *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
Schechtman M. 2014. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. New York, NY: Oxford University Press.
Schechtman M. 2020. Glad it happened: personal identity and ethical depth. *Journal of Consciousness Studies* **27**, 95–114.
Shoemaker D. 2019. Personal identity and ethics. In *Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), ed. E.N. Zalta. <https://plato.stanford.edu/archives/win2019/entries/identity-ethics/>.
Sugden R. 2004. The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. *American Economic Review* **94**, 1014–1033.
Sugden R. 2015. Looking for a psychology for the inner rational agent. *Social Theory and Practice* **41**, 579–598.
Sugden R. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. New York, NY: Oxford University Press.
Sunstein C.R. 2015. Nudges, agency, and abstraction: a reply to critics. *Review of Philosophy and Psychology* **6**, 511–529.
Sunstein C.R. 2019. *On Freedom*. Princeton, NJ: Princeton University Press.
Swinburne R. 1984. Personal identity: the dualist theory. In *Personal Identity*, ed. S. Sydney and R. Swinburne, 3–66. Oxford: Blackwell.
Taylor C. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
Thaler R.H. and H.M. Shefrin 1981. An economic theory of self-control. *Journal of Political Economy* **89**, 392–406.
Thaler R.H. and C.R. Sunstein 2003. Libertarian paternalism. *American Economic Review* **93**, 175–179.
Thaler R.H. and C.R. Sunstein 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Rev. and expanded edn). New York, NY: Penguin Books.
Williams B. 1970. The self and the future. *Philosophical Review* **79**, 161–180.

Ivan Mitrouchev is a postdoctoral researcher in economics at IESEG School of Management in Lille (France). His research interests belong to the methodological challenges of evaluating well-being when individuals do not conform to rational choice. URL: http://www.ivanmitrouchev.com.

Valerio Buonomo is a doctor in philosophy at the University of Milan (Italy) and a policy officer at the European Commission in Brussels (Belgium). His doctoral research has been devoted to the theories of personal identity over time, as well as other topics related to ontology and philosophy of time. Email: valerio.buonomo3@gmail.com. URL: https://www.centreforphilosophyoftime.it/valerio-buonomo/.

Cite this article: Mitrouchev I and Buonomo V (2024). Identity, ethics and behavioural welfare economics. *Economics and Philosophy* **40**, 310–336. https://doi.org/10.1017/S0266267123000123

https://doi.org/10.1017/S0266267123000123 Published online by Cambridge University Press