Identity, Personal Persistence and Normative Economics

Ivan Mitrouchev[†‡] and Valerio Buonomo[§]

[†]Univ Lyon, UJM Saint-Etienne, GATE UMR 5824, F-42023 Saint-Etienne, France
[‡]Université de Reims Champagne-Ardenne, REGARDS EA 6292, 51571 Reims Cedex, France
[§]Università degli Studi di Milano, Center for Philosophy of Time, Milano, Italia

revised and resubmitted for *Economics and Philosophy* (February 2021)[*]

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 the other. In this paper, we tackle this ethical question from the ontological question of personal persistence: *what does it take for an individual to persist from one time to another?* We review the main theories of personal persistence offered in analytic philosophy and discuss the philosophical problems related to the unified self assumption as an alternative to the multiple selves assumption in normative economics. We discuss two main issues. First, most of the authors defending a unified account of the self in normative economics tend to consider the question of identity over time from an *ethical* viewpoint but not from an *ontological* viewpoint. We argue that the ethical viewpoint is misleading because it reduces the question of personal persistence to the question of *personhood*. Second, we discuss the fact that the unified account of the self endorsed in the critical literature of behavioural welfare economics assumes the *narrative* view of personal persistence. Because of its many philosophical objections, we argue, however, that the narrative view cannot provide a satisfying account of identity.

Keywords. *behavioural welfare economics — identity — multiple selves — time preference — unified self*

JEL codes. B41, D15, I31

[*]We thank John Davis, Cyril Hédoin and Marya Schechtman for useful feedbacks on early drafts of this paper. All mistakes remain ours.

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).[1] 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 made reasonably (e.g. avoiding the temptation of eating a cake). 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 a recent 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?' (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?' (31)

[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 (see Kahneman (2011 [Ch.1]). The author makes it clear that system 1 and system 2 'are fictious characters ... [they] are not systems in the standard sense of entities with interacting aspects or parts' (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' (394), they introduce their model 'by viewing the individual as an organization' (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 known 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).

This philosophical problem is particularly salient when one makes the assumption of multiple selves. Indeed, ethical questions such as what the relationship between different selves are, or whether selves differ in the same ways as individuals differ seem to be unavoidable from a philosophical perspective. Yet the multiple selves assumption is not the only way to describe intertemporal choice. For various reasons some authors are sceptical about the multiple selves assumption and instead propose an alternative assumption of the unified self, in which the individual is represented as a *unified* being through time.

Among these authors, Sugden (2004) drops the multiple selves assumption in his reconstruction of welfare economics without the concept of preference by considering instead the individual as 'a continuing locus of responsibility ... to the extent that, at each moment in her life, she identifies with her own actions, past, present, and future' (1018). In his theory of identity in economics, Davis (2011) criticises the fragmentation and dissolution of the individual into multiple selves and argues that a criterion of identity should satisfy what he calls the *individuation* and *reidentification* criteria (to be discussed below). The author argues that a unified concept of the self based on a capability approach (Nussbaum and Sen 1993) best satisfies his two criteria. In a similar trend, Hédoin (2015) has a narrative view of identity based on Korsgaard's (1989) definition of agency, which he argues, palliates the problem that selves cannot be attributed moral responsibility. Furthermore, the philosophical issues related to identity in behavioural welfare economics lead Dold and Schubert (2018) to suggest that 'the dualistic concepts of the individual should be abandoned in favor of a notion of a unified self that is constituted by its capacity to learn and reflect upon new preferences on a continuous basis' (221). Although appealing, the main problem with most of these alternative views of the unified self is that they are based on ethical assumptions rather than on *ontological* theories about *what makes an individual truly one*.[3] But arguing how an individual remains the same from one time to another inevitably results in debating about the ontological status of *what it takes for an individual to persist over time*.

The aim of this article is to clarify the philosophical difficulties of endorsing the assumption of the unified self as an alternative to the multiple selves assumption in normative economics. To our knowledge, these difficulties have not been investigated yet. The literature of identity-and-economics is, of course, extensive and goes beyond the philosophical problems of behavioural welfare economics (Akerlof and Kranton 2000, 2010; Davis 2003; Kirman and Teschl 2004). We do not, however, exclude the possibility for our study to be relevant to the broader program of identity-and-economics, nor to normative economics in general. By behavioural welfare economics (BWE) we refer to the literature, which reinterprets normative economics when individuals have incoherent preferences.[4] We include in this literature Camerer et al. (2003), Thaler and Sunstein (2003, 2009), Bernheim and Rangel (2007, 2008, 2009) and Bernheim (2009, 2016).[5] The ethical problems of BWE related to identity have particularly been a matter

[3]An exception is Davis (2011), to be discussed below.

[4]Here we avoid focusing on why preferences can be incoherent (e.g. present bias, framing effect or loss aversion), 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.

[5]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

of interest for Ferey (2011), Lecouteux (2015) and Hédoin (2015) in relation to Parfit's (1984) theory of identity. Although we have no particular objection to the philosophical issues of the multiple selves assumption raised by these authors, we argue that making the unified self assumption is nonetheless no less problematic as it requires economists to justify the underlying personal persistence theory on which their conception of identity is based. We introduce the literature of personal persistence into this 'multiple selves *versus* unified self' debate in normative economics which, we argue, is an important matter of interest for two reasons.

First, personal persistence provides what we judge to be a better framework for discussing the relationship between identity and ethics. The point is that many of the alternative assumptions of the unified self proposed in the literature do not take an *ontological* but an *ethical* viewpoint of the criterion of identity. To make sense of the distinction between these two viewpoints, consider the following example. If John changes his preferences over time (e.g. he wishes he had not eaten the cake yesterday), it is not clear which of his possible conflicting preferences over time should be considered as 'more important to him'. The ethical viewpoint would try to define the concept of personhood (typically an individual endowed with some psychological properties) and then, on the assumption that John is a person, would try to make sense of the identity of John over time (e.g. John persists over time because he is a person who has the ability of thinking about his past and future actions and then making sense of all his actions). According to the ethical viewpoint of identity, it is the moral authority of some selves (or one self) of an individual over his other selves that defines personal identity over time. In contrast, the ontological viewpoint would try to find out what makes John persisting from one time to another *and then* to see what the ethical implications for John could be (e.g. whether it makes sense that the preferences of John today have more importance than the preferences of John yesterday). Our position is that 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. But 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*, does it make sense to ground normative authority on some specific features (e.g. psychological ones). As things stand, we argue that the ethical viewpoint of identity is methodologically misleading since the question of identity over time cannot be reduced to the question of personhood ('what does it take for something to be a person?').[6] In fact, solving ethical problems of identity in normative economics

program, which aims at 'reconciling normative and behavioural economics' in response to the observation that individuals have incoherent preferences, see McQuillin and Sugden (2012).

[6]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

does not necessarily require making *ethical* judgements about the constitution of personhood. Instead, one has good reason to proceed in a different way, focusing first of all on the ontological criterion of identity over time, which is the basis for the way persons persist.

Second, the literature of personal persistence — as yet unknown to the literature of identity-and-economics — enlightens the difficulties associated with the current alternative assumptions of the unified self proposed in the critical literature of BWE. More particularly, some economists who are sceptical about the multiple selves assumption endorse the *narrative* view of the unified self in order to avoid the philosophical problems associated with the multiple selves assumption (Sugden 2004, 2018; Davis 2011; Hédoin 2015; Gallois and Hédoin 2017; Dold and Schubert 2018). We however bring about some philosophical arguments that give us good reason to be also sceptical about the narrative view of identity.

The rest of the paper is organised as follows. Section 2 introduces an example of the identity problem in ethics applied to BWE. Section 3 contrasts our ontological approach of identity with related literature. Section 4 presents the framework of personal persistence by formalising the criterion of identity over time. Section 5 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 criticisable than the multiple selves assumption because (i) they tend to define identity from an ethical viewpoint and (ii) because the narrative view is philosophically problematic. Section 6 concludes and briefly discusses the relevance of our 'ontological-first' approach for normative economics, paving the way for further research.

Two useful clarifications should be stated early. First, 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.[7] Second, *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,

personhood, persistence and others), see Olson (2020). We come back to the relation between persistence and personhood questions in Sections 4 and 5.

[7] 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 to BWE, see Ferey (2011, 746-747), Lecouteux (2015, 403-407) and Hédoin (2015, 98-102).

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 Sunstein (2019) and Kahneman (1994) 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 see how it can provide a better viewpoint compared to the related literature, 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 normative rules for individual decision-making, but also to critically assess the 'unified self' alternatives to such an ethical problem.

Assuming the view of multiple selves, an individual I can be considered over time as a collection of a finite number of temporal selves $\{s_1, ..., s_n\}$, where s_i is a temporal self which exists at a given time t_i. Let $i = \{1, ..., 100\}$ 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.[8] As shared by Sunstein (2019) and Kahneman (1994), one central matter of concern in intertemporal choice is which one of the two (or of the many other) selves has/have normative authority.[9] 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, particularly when these rules are based on evidence. This brings us to the formulation of the ethical problem of identity in the ontological framework of personal persistence, which we introduce 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 of her present temporal parts. 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 economics, where each s_i is 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) view that economists should rather promote institutional arrangements so that individuals can seek what they want — disregarding how incoherent their preferences are.

[8]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).

[9]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 such self will actually be John (and not somebody else). For a presentation of the relationship between personal persistence and ethics, see Shoemaker (2019).

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 rule would however be endorsed by BWE only if it appears that s_1 is the far-sighted planner, cold state or system 2. But 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 viewpoint, particularly when we have no objective criterion to determine what makes the normative authority of s_1 more important than s_{100} (and conversely). As Sunstein (2019) 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' (79).

2.3 The Objective Rule

Another rule would then 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 their past, present and future status. There may be at least two ways to define such 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, $\{s_1, ..., s_{51}\}$ but not $\{s_{52}, ..., s_{100}\}$ — then fifty-one selves against forty-nine overrule whatever s_{100} states, and then the preference of A over B should be taken as the one which is normatively relevant. This view 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 fifty-one selves are the myopic doers and the forty-nine 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) recognise 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 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 all the selves is the 'supreme' one. For example, assume only s_{30} and s_{100} prefer B to A (among all the remaining selves), and we discover from certain knowledge that s_{30} has supreme normative authority. Using this objective rule, we could therefore account for the preference of s_{30}. This view is endorsed by Bernheim and Rangel (2007, 2008, 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 here important: some temporal selves could be eliminated from the possibility of having any normative authority — typically those 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 true preference, e.g. all temporal selves included in the set $[s_{30}, ..., s_{40}]$, and assume that we could somehow determine such 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 reinterpret normative economics when individuals have incoherent preferences. This means that if one self which belongs to the set $[s_{30}, ..., s_{40}]$ 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 Viewpoint on 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) justifiably underlines the problem of 'heavily value-laden language' (38-39), 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 emphasises 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 maximises 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 (89). This social choice alternative is also well recognised by Sunstein (2019), who notes that

> '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.' (76)

Aggregating well-being over time would however ultimately yield to the arbitrary ethical rules previously discussed.[10] 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 from a *personal persistence* perspective (what makes an individual *one* over time). Although we are not at all against the social choice alternative, we take the ethical issue of identity at its roots by focusing on the *criterion of identity over time*. The way we see the ethical identity problem 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 receivable 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.[11] 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 related 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 on what makes the individual persist from one time to another. This means we aim to take the reverse stance of Sugden (2004), Hédoin (2015) and Dold and Schubert (2018) in their account of identity. We first ask what makes an individual persist over time *and then* this ontological enquiry will help us comment on the ethical problems faced in normative economics. 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 related literature, we do not say anything on what is required to have moral responsibility.[12]

[10]We of course do not mean these rules to be exhaustive. Any other rule is receivable, for it has reasons to be justified.

[11]We say 'receivable' because some philosophers object to the view that any ontological or metaphysical question about personal persistence is relevant to our moral practical concerns (Rovane 1998; Conee 1999; Johnson 2017).

[12]In any case, one may simply argue that to take a position on what it takes for a living entity to have

In addition, it would be in our view misleading to seek evidence if one wants to define personal identity. Evidence would provide what is required to reasonably claim that a person persists over time, e.g. observing the physical properties of an individual over time. But evidence is in fact neither necessary for personal identity over time — one may change some physical properties and/or forget some of her past events, but nonetheless persist over time — nor sufficient for that — we may imagine the case in which there is another person, who resembles a friend of ours, and is aware of the most important events of her life. In other terms, evidence provides answers to the so-called evidence question 'how do we find out whether a person at one time is numerically identical to a person at another time?' (i.e. 'what evidence do we need to maintain that a person we see today is or is not the same person we saw yesterday?'), but not to the ontological question 'what are the constitutive conditions of personal identity over time?' (i.e. 'what does it take for a person to persist from one time to another?').

In order to avoid (i) intuitive reasoning that relies on common sense about which self has normative authority; (ii) a social choice approach which would consist in proposing ethical rules to aggregate intrapersonal well-being, and (iii) the evidential limits of defining personal identity, we propose to formulate the identity problem of BWE within the framework of personal persistence from an ontological viewpoint. Contrary to the evidence viewpoint, we ask ourselves what the constitutive conditions of personal identity over time are instead of what the factual evidence is to conclude that a person persists over time. In the next section, we introduce such framework (Buonomo 2018) and then discuss in Section 5 four main theories that aim at explaining the unification of an individual's temporal selves.

4 The Criterion of Identity over Time

The problem of personal persistence consists in focusing on the criteria of identity over time. 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.
Let also $Px \vee Py$, where P is the property of being a person.[13]
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.

moral properties is a question that belongs to metaethics. 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?').

[13] Such criterion of identity over time seems to violate Leibniz's indiscernibility of identicals, which states that $x = y \implies \forall F(Fx \iff Fy)$. 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 'xIy' to denote the identity relation, as '$xIy \iff x = y$' does not have to hold if we offer alternative interpretations of the I-relation (we thank one anonymous referee for pointing out this to us). Although we are very much sympathetic to this remark, asking for Leibniz's indiscernibility of identicals for identity over time appears too strict — a 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

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 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.' (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.' (99)

These conceptions of identity already 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 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 (ii) the question of the concept of personhood ('what does it take for something to be a person?') or even (iii) the question about the biographical identity of persons ('who am I?').[14]

We recognise 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 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. We particularly think of the following identity conditions proposed by Hédoin (2015) based on Korsgaard's (1989) representation of agency.

properties. For instance, consider John being ten years old and John being forty years old. The physical and mental properties of ten-years-old-John and forty-years-old-John are likely to be different (e.g. they have different heights, ways of thinking, etc.). But if we take Leibniz's indiscernibility of identicals to be true, ten-years-old-John and forty-years-old-John cannot be identical. In short, Leibniz's indiscernibility of identicals seems reasonable for *synchronic* identity, but not for *diachronic* 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'.

- *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 the next section we review and discuss the main theories of personal persistence offered in the literature of analytic philosophy in the light of the framework we have just introduced. The presentation of these theories nurtures the 'multiple selves *versus* unified self' debate in normative economics in the way that it exposes several possible views of the unified self, which are not necessarily narrative. Importantly, knowing that the narrative view is dominant in the critical literature of BWE, we discuss few of its philosophical problems. Because of the objections associated with most of the theories of personal persistence we present, we end up suggesting that economists interested in the ethical identity problem of BWE should better allocate their efforts in making correct assumptions about the ontology of individuals rather than trying to make correct assumptions about personhood.

5 Theories of Personal Persistence[15]

5.1 The Psychological View

The psychological view claims that an individual is identical over time in 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.

[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 normative 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.

Let x be an entity that exists at time t_i and y an entity that exists at time t_j.
Let also $Px \lor Py$, 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 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 preference, which is an essential property of the 'inner rational agent' (Infante, Lecouteux, and Sugden 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 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 this economic agent cannot make way for other 'psychological roles' when she makes a decision, e.g. making a decision as a parent or a wife.[17] It is

[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 characterised by the (very non-standard) rejection of the assimilation between 'personal identity' and 'personal persistence' — commonly summarised in Parfit's famous sentence 'identity does not matter for survival'. Given this account, the ethical claim of personal identity is prioritised, whereas the ontological question follows. 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 generalisable argument to face the first objection against psychological views. On the revisionary aspects of Parfit's theory, see Rovane (1998, 11) and Martin (1998, 15).

[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. If seriously claimed, we believe this point to be misleading, considering that leading behavioural economists 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 examples to justify libertarian paternalism, such as avoiding the temptation of eating too much of the cashew bowl nuts

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, such psychological view may miss something that the critical authors of BWE have argued to be important for identity: the *meaning* one attributes to her 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 cognitive 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 (an assumption we are so far quite sympathetic to).

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.
> 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

before dinner (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 from any kind of life decision (such as not eating cashew nuts before dinner or using a GPS), we seriously assume the view that economic theory can explain any kind of choice that involves intertemporality.

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

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 seems 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). In virtue of biological continuity, advocates of the physical view would argue that the cerebrum-less donor remains the same person, while 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 emphasise 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] In contrast, the narrative view introduced below is certainly the one which has the most tenants in the critical literature of BWE (Sugden 2004, 2018; Hédoin 2015; Gallois and Hédoin 2017 and Dold and Schubert 2018). We will thus spend more time discussing its philosophical issues.

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.

[19] In response to Lecouteux's (2015) criticism, according to which libertarian paternalism presents an implausible model of identity, Sunstein (2015) mentions the possibility of considering the physical view as an alternative to Parfit's reductionist account of identity. In his words, 'Consider a competing 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' (527 — his emphasis). We do not however believe that this point is raised seriously by Sunstein (at least not in 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).

Let x be an entity that exists at time t_i and y an entity that exists at time 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 *characterisation* 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 prioritise the characterisation 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 (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; 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 viewpoint, i.e. independently from the first-person viewpoint. 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

[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 (discussed below).

those for which the agent 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' (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 characterise moral accountability than a narrative one.

Second, the narrative view endorsed in the critical literature of BWE leads to the following disturbing paradox. Recall that authors who reject the assumption of multiple selves also reject the idea that a far-sighted planner exists in 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 the case 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 characterisation 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) puts it, 'the identity of that subject of experiences must be preserved across time for its experiences to be so gathered up' (sec. 2.3.). 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 in virtue of being persons. The concept of person, broadly defined, is an individual who has the ability of being morally responsible. Thus, identity is reduced to an individual who has moral thoughts, and the question of personhood would then require an answer regarding what makes 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 psychological continuity, persists through time (like 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. For these reasons, we are inclined to reject the narrative view of personal persistence as it presupposes the identity it is called 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.

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 should not be taken as a basic condition of identity. 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).

In addition to giving importance to the numerical condition over the characterisation condition, we have just argued that the narrative view is philosophically problematic. Is it then all what theories of personal persistence have to offer to normative economics? Before concluding, we would like to mention another unified view that may overcome some of the methodological problems of the three views previously discussed. In particular, we briefly discuss the theory of identity of Davis (2011), which is a sort of hybrid theory between the narrative view and the sociological view (that we now introduce).

[21] If the sceptical reader wishes to criticise our present result, we invite him/her to dispute our three informative conditions of the criterion of identity over time presented in Section 4. 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).

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* viewpoint. But for each of these views, neither the social status of identity — how individuals are contextualised in their social environment — nor the story of their life told from a *third-person* viewpoint is suggested. The sociological view can instead be formulated as follows.

$$\begin{aligned}
&\text{Let } x \text{ be an entity that exists at time } t_i \text{ and } y \text{ an entity that exists at time } t_j.\\
&\text{Let also } Px \vee Py, \text{ where } P \text{ is the property of being a person.}\\
&\text{Then } x = y \text{ if and only if } x \text{ and } y \text{ are connected by some sociological relations.}
\end{aligned}$$

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 then is what constitutes persons over time. In effect, if social systems constitute personhood, then this view can be informative on how individuals' normative authority can be determined.

According to Schechtman (2014 [Ch. 5]), human beings are not only characterised in virtue of their biological and psychological features, but also in 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 characterises an individual. That is to say, every social factor that characterises a human being born in a given environment (her culture, habits, norms) is her ontological unit that gradually becomes responsible and concerned for its own future (Shoemaker 2019). Such responsible unit is then no different from the embryo from which she evolved, and then 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 the one of Davis (2011). Davis provides an extensive account of 'socially embedded individuals' [Ch. 3] but in contrast to 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.' (213)

Davis particularly criticises the model of social identity of Horst, Kirman, and Teschl (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 characterises through the notion of capability (Nussbaum and

[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 over the two former.

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' (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' (36) than by individuals making choices in relative isolation.[24]

We think the sociological view may be an interesting alternative to the ethical problem of identity we have so far discussed, especially when a consequent body of empirical studies supports the view that individual 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 viewpoint. 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 viewpoint?' (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 a common practice presented 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 healthy, 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.

[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.

[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 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]).

6 Concluding Remarks

In the present paper, we aim to show that the assumption of the unified self in the critical literature of BWE is philosophically problematic when studied through the lens of personal persistence. We introduce the framework of personal persistence (Buonomo 2018) in order to discuss the alternative views of the unified self economists may endorse as responses to the multiple selves assumption in BWE. We emphasise that most of the unified views presented above (especially the *narrative* one) are philosophically problematic, and that the identity criterion — even when discussed from the viewpoint of ethics — is better defined by a numerical instead of a characterisation condition. We stress that our 'ontology-first' viewpoint of identity does not only show the philosophical difficulties of assuming the unified view of the self in normative economics. It also offers new directions of research for those who are convinced that the question of personhood cannot be reduced to the question of personal persistence, and therefore that any question about the persistence of an individual cannot escape from the ethical problem of identity with purely 'ethical' considerations about what a person is. We advance two clarifications that we hope will stimulate perspectives of future research.

First, focusing on the ontological question of personal persistence does not set aside the ethical concerns one can have within BWE, and more generally within normative economics. On the contrary, ethical concerns about identity — such as those discussed by Sunstein (2019) and Kahneman (1994) — can be informed by the ontological account of identity. If it appears that John can be unified through time with some relations (either psychological, physical, narrative or sociological), then it can make sense to attribute a meaning to all of his actions over time. But if ten-years-old-John and forty-years-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. As we have stated it, 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 our efforts in finding a better account of the unified self (e.g. the sociological view) *before even discussing the ethical rules of identity over time*. In this sense, our 'ontology-first' approach can support and eventually steer some discussions within the identity debate in normative economics.

Second, considering the ethical problem of identity through the lens of the ontological viewpoint does not mean building or defending a theory of personal identity *per se*. Our message is that economists should not worry too much about the personhood question because the criterion of identity over time can be discussed without making any ethical claim. This is true even if their practical worries are ethical. Indeed, in order to provide an answer to the ethical dilemmas such as the ones raised by Sunstein (2019) and Kahneman (1994), one has deep interest in first asking what makes John actually the same from one time to another, so that it makes sense to give some normative relevance to his actions through time. That is, it is the *result* of how individuals persist through time that has consequences on our ethical concerns about identity and not *vice versa*, which is what happens when we inquire 'which theory of personal persistence best fits with the way we want to represent our idealistic picture of the morally responsible individual?'.[26]

[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

Many points remain, of course, unexplored, such as the possibility of combining the narrative view with the sociological view, or a detailed assessment of whether the 'narrative-sociological' view is the most promising theory of personal persistence. With our present analysis, we hope to have provided economists a comprehensive approach to the philosophical objections of the unified self, that are not only important for the future of behavioural welfare economics, but more generally for the future of normative economics.

References

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

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Akerlof, G. A. and R. E. Kranton (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-being*. Princeton University Press.

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 I. Brocas and J. Carrillo (Eds.), *The Psychology of Economic Decisions. Volume One: Rationality and Well-Being*, pp. 137–67. Oxford University Press.

Benjamin, D. J., J. J. Choi, and A. J. Strickland (2010). Social identity and preferences. *American Economic Review 100*(4), 1913–28.

Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association 7*(2-3), 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*(1), 12–68.

Bernheim, B. D. and A. Rangel (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review 97*(2), 464–470.

Bernheim, B. D. and A. Rangel (2008). Choice-theoretic foundations for behavioral welfare economics. In A. Caplin and A. Schotter (Eds.), *The Foundations of Positive and Normative Economics*, pp. 155–192. Oxford University Press.

Bernheim, B. D. and A. Rangel (2009). Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics 124*(1), 51–104.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2008). How are preferences revealed? *Journal of Public Economics 92*(8-9), 1787–1794.

Blatti, S. and P. F. Snowdon (Eds.) (2016). *Animalism: New Essays on Persons, Animals, and Identity*. Oxford University Press.

view is the best account we can find of personal persistence, ethical claims which would derive from this personal persistence view is another philosophical question to be solved. Although we are well aware of this potential objection, a metaethical assessment of Hume's 'is-ought' problem is far out of the scope of the present paper.

Buonomo, V. (2018). A brief guide to personal persistence. In V. Buonomo (Ed.), *The Persistence of Persons: Studies in the Metaphysics of Personal Identity over Time*, pp. 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*(3), 1211–1254.

Chen, Y. and S. X. Li (2009). Group identity and social preferences. *American Economic Review 99*(1), 431–457.

Conee, E. (1999). Metaphysics and the morality of abortion. *Mind 108*(432), 619–646.

Davis, J. B. (2003). *The Theory of the Individual in Economics: Identity and Value*. Routledge.

Davis, J. B. (2011). *Individuals and Identity in Economics*. Cambridge University Press.

DeGrazia, D. (2005). *Human Identity and Bioethics*. Cambridge University Press.

Dold, M. F. and C. Schubert (2018). Toward a behavioral foundation of normative economics. *Review of Behavioral Economics 5*(3-4), 221–241.

Ferey, S. (2011). Paternalisme libéral et pluralité du moi (Libertarian paternalism and multiple-selves). *Revue Économique 62*(4), 737–750.

Gallois, F. and C. Hédoin (2017). From identity to agency in positive and normative economics. In *Forum for Social Economics*, pp. 1–17. Taylor & Francis.

Gul, F. and W. Pesendorfer (2001). Temptation and self-control. *Econometrica 69*(6), 1403–1435.

Hédoin, C. (2015). From utilitarianism to paternalism: when behavioral economics meets moral philosophy. *Revue de Philosophie Économique 16*(2), 73–106.

Hershenov, D. (2005). Do dead bodies pose a problem for biological approaches to personal identity? *Mind 114*(453), 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), 1–25.

Johnson, M. (2017). The personite problem: should practical reason be tabled? *Noûs 51*, 617–44.

Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft 150*(1), 18–36.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin Books.

Kirman, A. and M. Teschl (2004). On the emergence of economic identity. *Revue de Philosophie Économique 9*(1), 59–86.

Korsgaard, C. M. (1989). Personal identity and the unity of agency: a Kantian response to Parfit. *Philosophy & Public Affairs 18*(2), 101–132.

Köszegi, B. and M. Rabin (2007). Mistakes in choice-based welfare analysis. *American Economic Review 97*(2), 477–481.

Köszegi, B. and M. Rabin (2008). Revealed mistakes and revealed preferences. In A. Caplin and A. Schotter (Eds.), *The Foundations of Positive and Normative Economics*, pp. 193–209. Oxford University Press.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics 112*(2), 443–478.

Lecouteux, G. (2015). In search of lost nudges. *Review of Philosophy and Psychology 6*(3), 397–408.

Lecouteux, G. (2016). From homo economicus to homo psychologicus: the Paretian foundations of behavioural paternalism. *Œconomia. History, Methodology, Philosophy 6*(2), 175–200.

Loewenstein, G. and R. H. Thaler (1989). Anomalies: intertemporal choice. *Journal of Economic Perspectives 3*(4), 181–193.

Lowe, E. J. (2012). The probable simplicity of personal identity. In G. Gasser and M. Stefan (Eds.), *Personal Identity: Complex or Simple?*, pp. 137–155. Cambridge University Press.

MacIntyre, A. (1984). *After Virtue*. University of Notre Dame Press.

MacIntyre, A. (1989). The virtues, the unity of a human life and the concept of a tradition. In S. Hauerwas and L. G. Jones (Eds.), *Why Narrative?* W.B. Eerdmans.

Martin, R. (1998). *Self-Concern: An Experiential Approach to What Matters in Survival*. Cambridge University Press.

McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford University Press.

McQuillin, B. and R. Sugden (2012). Reconciling normative and behavioural economics: the problems to be solved. *Social Choice and Welfare 38*(4), 553–567.

Noonan, H. W. (1998). Animalism versus Lockeanism: a current controversy. *The Philosophical Quarterly 48*(192), 302–318.

Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

Nussbaum, M. C. and A. Sen (Eds.) (1993). *The Quality of Life*. Oxford University Press.

Olson, E. (1997). *The Human Animal: Personal Identity without Psychology*. Oxford University Press.

Olson, E. T. (2003). An argument for animalism. In R. Martin and J. Barresi (Eds.), *Personal Identity*, pp. 318–334. Blackwell.

Olson, E. T. (2020). Personal identity. *The Stanford Encyclopedia of Philosophy*.

Olson, E. T. and K. Witt (2019). Narrative and persistence. *Canadian Journal of Philosophy 49*(3), 419–434.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

Ross, D. (2005). *Economic Theory and Cognitive Science: Microexplanation*. MIT Press.

Ross, D. (2014). *Philosophy of Economics*. Palgrave Macmillan.

Rovane, C. (1998). *The Bounds of Agency*. Princeton University Press.

Schechtman, M. (1996). *The Constitution of Selves*. Cornell University Press.

Schechtman, M. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford University Press.

Shoemaker, D. (2019). Personal identity and ethics. *The Stanford Encyclopedia of Philosophy*.

Sugden, R. (2004). The opportunity criterion: consumer sovereignty without the assumption of coherent preferences. *American economic review 94*(4), 1014–1033.

Sugden, R. (2015). Looking for a psychology for the inner rational agent. *Social Theory and Practice 41*(4), 579–598.

Sugden, R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press.

Sunstein, C. R. (2015). Nudges, agency, and abstraction: a reply to critics. *Review of Philosophy and Psychology 6*(3), 511–529.

Sunstein, C. R. (2019). *On Freedom*. Princeton University Press.

Swinburne, R. (1984). Personal identity: the dualist theory. In S. Sydney and R. Swinburne (Eds.), *Personal Identity*, pp. 3–66. Blackwell.

Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Harvard University Press.

Thaler, R. H. and H. M. Shefrin (1981). An economic theory of self-control. *Journal of Political Economy 89*(2), 392–406.

Thaler, R. H. and C. R. Sunstein (2003). Libertarian paternalism. *American Economic Review 93*(2), 175–179.

Thaler, R. H. and C. R. Sunstein (2009). *Nudge: Improving Decisions about Health, Wealth, and Happiness* (revised and expanded ed.). Penguin Books.

Williams, B. (1970). The self and the future. *The Philosophical Review 79*(2), 161–180.