Philosophical Challenges of Behavioural Public Policy*

Preprint version of Giancarlo Ianulardo, John Davis, and Ricardo Crespo (Eds). *The Elgar Handbook of Teaching Philosophy to Economists*. Edward Elgar.

Ivan Mitrouchev†

This version: 16/06/2024

Abstract

If evidence shows that people are not rational, how to know what makes them better off? Over the past few decades, there has been a surge of interest among economists on how to interpret "irrational" behaviour for welfare evaluation. This interest has been intensified in the public sphere, with the institutions of behavioural public policies all over the world. In this chapter I discuss two philosophical challenges of behavioural public policies. To whom should behavioural public policies be addressed? How to evaluate individual welfare when choices are inconsistent over time?

*I am grateful to Valerio Buonomo, John Davis, Guilhem Lecouteux and Robert Sugden, who have largely influenced me regarding the questions I address here. All mistakes remain my own.

†IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, F-59000 Lille, France.

1 From Behavioural Economics to Behavioural Public Policy

Standard models of decision making assume several rules (or axioms) of human behaviour, which many have – at some point – questioned their empirical realism, usefulness, or both. Because those axioms are theoretical constructs, they may not be observed in the real world. For example, in many cases people may not know what they want (violation of *completeness*), they may have cyclical preferences (violation of *transitivity*), or they may be influenced by the way a choice problem is formulated (violation of *invariance*). During the 1970s, the heuristics-and-bias programme was devoted to test whether standard models of decision making – more particularly expected utility theory – conform to real choices of individuals, and if not, to what extent actual choice diverges from the axioms of rational choice, as embodied in standard models of decision making (Tversky and Kahneman 1973, 1974; Kahneman and Tversky 1979).

But from the 1990s, some researchers became interested in another question: how to interpret rationality deviations for *normative* decision making?[1] Should they be interpreted as errors of *reasoning* – which rather refers to a *cognitive* interpretation or rationality deviations – or should they be interpreted as failures to be *better off* – which rather refers to an *ethical* interpretation of rationality deviations? Different approaches aimed to address this question. While some suggested that cognitive biases constitute evidence that people fail to choose according to what makes them better off (Kahneman, Wakker, and Sarin 1997), other uphold both (*cognitive* and *ethical*) interpretations (Thaler and Sunstein 2003), while other proposed alternative models of normative decision making by holding minimal assumptions on what makes people better off (Sugden 2004; Bernheim and Rangel 2007, 2009).[2]

Although this concern was initially purely academic, the challenges regarding "irrational behaviour" for normative decision-making did not delay the use of behavioural insights in public policy. This is mainly due to the international success of Thaler and Sunstein (2003)'s libertarian paternalism with their seminal and highly influential *Nudge* (Thaler and Sunstein 2009). Nudging took shape in real-world policies with (among others) the institution of the Behavioural Insights Team in 2010 by the government of the United Kingdom, as well as the Foresight and Behavioural Insights Unit in 2014 by the European Commission.[3] Behavioural Public Policy (BPP) can be defined as the application of behavioural insights to the design and implementation of public policies. In this sense, it is evidence-based – i.e. it relies on the academic literature of individual decision making, which is based on lab and field *experiments*.

This chapter aims to discuss two philosophical challenges of BPP.[4] The first is about the status of the policymaker: *to whom* should BPP be addressed? The second is about inconsistent behaviour over time. Because this phenomenon is now well documented (with hyperbolic or quasi hyperbolic discounting models, providing a better description of intertemporal choice), an important question for the policymaker is to know which preferences over time should count for individual welfare.

[1]One may object that before even asking how to interpret rationality deviations for normative decision making, economists should, first of all, agree on what is meant by *rationality*. This is far from an easy task, as no definition is consensual. According to Wakker (2010), to be rational is to be an *expected utility maximiser*. According to Gilboa and Schmeidler (2001), to be rational is *not to regret one's choice* when one becomes aware of the consequences of his/her choice. According to Todd and Gigerenzer (2012), rationality is *ecological* in the degree that human heuristics are often well adapted to the structure of the environment. In this chapter, I take a narrow definition of rationality: a behaviour is rational when it conforms to some principles/rules (whatever those principles/rules are, e.g. *completeness*, *transitivity*, *invariance*, among many others).

[2]See Mitrouchev (2024) for a literature review on how the heuristics-and-biases programme has branched out around this normative issue.

[3]See Halpern (2015) for the history of how nudging found its way into public policy from his own perspective – as the President and Founding Director of The Behavioural Insights Team.

[4]There are, of course, many issues other than those addressed in the present chapter. See e.g. Oliver (2013), among others.

2 To Whom Should Behavioural Public Policy Be Addressed?

2.1 Paternalism: The Standpoint of the Policymaker

The original proposition of Thaler and Sunstein (2003) can be seen as questioning the *liberal* tradition of economic policy. In the behavioural paradigm, because it is no more assumed that people make "rational" choices, and because rational choice used to be the standard of welfare in economic policy, is it legitimate for a policymaker to *interfere* with people's choices? In other words, under the assumption that rationality still constitutes/indicates individual welfare, is *paternalism* justified, and can the cost of violating people's freedom to choose whatever they want be avoided?

Emphasising "the possibility that in some cases individuals make inferior choices, choices that they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower" (Thaler and Sunstein 2003, 175), the authors' proposition can be interpreted as an extension of Pareto efficiency to individual psychology. That is, if a social situation can be improved by "debiasing" some individuals (thus increasing their own welfare) without decreasing the welfare of others (who may be fully rational), then such a social situation is to be preferred over another. A similar form of paternalism was advocated by Camerer et al. (2003) in terms of *interpersonal comparisons of utility*. In their definition of *asymmetric* paternalism, "a policy is asymmetrically paternalistic if it creates large benefits for those people who are boundedly rational [. . .] while imposing little or no harm on those who are fully rational" (1219).

Both libertarian and asymmetric paternalisms are embedded in the long strand of welfare economics, which evaluates social situations from an *external standpoint*: the *policymaker* (or "social planner") – a benevolent individual whose aim is to maximise individuals' utilities (and therefore social welfare).[5] These forms of paternalism can be categorised as "welfarist" (Sen 1979), in the sense that they exclusively take *personal utility* as the informational basis for evaluating a social situation. That is to say, *liberty* – which is a value that is given an important role in *libertarian* and asymmetric paternalisms – is (paradoxically) never given full consideration, as its definition and measure is outside the scope of these propositions.

This poses a problem for BPP for at least two reasons: (i) the compatibility between welfare and liberty is far from self-evident (Grüne-Yanoff 2012; Scoccia 2019) and if so, (ii) it is not clear how people can make trade-offs between welfare and liberty. If we look for an alternative approach to paternalism, we need to stand aside from the *policymaker* standpoint and *welfarism*, and to take a closer look to the *contractarian* and *liberal* traditions in political economy.

2.2 Anti-Paternalism: The Standpoint of Individuals

This is the path taken by Sugden (2004, 2018), who argues that inconsistent behaviour is, by no means, a reason to believe that people are necessarily worse off. Instead, he proposes to evaluate social situations according to an *opportunity* criterion, which states that it is good to enhance people's opportunity sets to choose whatever they want, no matter if their choices are rational or not.

Sugden's approach is Smithian in the way that it is (as the author emphasises it) *a defence of the market*. In his philosophy, the market is not to be seen as an institution of competition but of cooperation. This representation of the market is influenced by his reading of Smith (1759), where he explains economic behaviour through self-love motivations instead of self-interest. Empirically, this behaviour can be observed in trust game experiments, where subjects have a tendency to behave reciprocally towards the other(s). The main implication is that there is no need to intervene in a market economy where individuals naturally

[5] This is permitted with the Pareto criterion in Thaler and Sunstein (2003) or with cost-benefit analysis in Camerer et al. (2003). In BPP, the policymaker is no more an abstraction: it is a team of individuals working on how to effectively design a policy for changing behaviour in many domains such as health, education or finance (among others).

aim to increase their mutual benefit through voluntary exchanges. His approach is Millian in the way that it only justifies paternalism when it prevents harm to others, but not when it prevents failures of well-being. In Sugden's reading of Mill (1859), (i) cooperation for mutual benefit is a governing principle of social life, (ii) competitive market is a network of mutually beneficial transactions, and (iii) in cooperative relationships, it is for each individual to judge what counts as his/her benefit. It is then Buchananian in the way that economics (including economic policy) is not about *rational choice* but about *voluntary exchange* (Buchanan 1964). This implies that social arrangements are not assessed from the viewpoint of a benevolent social planner but from the viewpoints of the individual members of a society (Buchanan 1987). This is the *contractarian* approach he defends. Contrary to the social planner in libertarian and asymmetric paternalisms, and the policymaker in BPP, whose aim are to maximise welfare, a contractarian theorist asks whether it is in the interest of each individual to accept the rules of that institution, on the condition that everyone else does the same.

In Sugden's anti-paternalist approach, individuals' situations are then not to be evaluated from an *external* standpoint (the social planner) but from an *internal* standpoint (individuals themselves). In addition, because nobody is best judge than oneself to know what counts as his/her personal well-being, there is no need for a *welfare* criterion. The only criterion needed is one that can assess how much opportunities have been exhausted in voluntary transactions (see Sugden 2018, Ch. 5).

2.3 Towards a Third Alternative? The Standpoint of "Enlighted" Individuals

As previously mentioned, "irrational" behaviour is essentially a matter of *interpretation*. This means there is no straightforward answer about how to interpret "irrational behaviour" because any answer would lead us to value judgements (or perhaps more precisely, ethical beliefs) about what *should be*.[6] Because value judgements are complex and plural, economists can hardly reach a consensus on how to interpret rationality deviations. Each economist may hold a particular vision of (i) whether biases are actually "biases", (ii) how to interpret these biases, (iii) whether to be paternalistic or not, and perhaps more generally, (iv) what economic policy is about. We now better understand to what extent the question of how to evaluate welfare when people deviate from rational choice has carried us towards the question of the *addressee* of BPP – i.e. *to whom should BPP be addressed?* Should it be addressed to a *policymaker* whose aim is to make people better off (Thaler and Sunstein 2003; Camerer et al. 2003), should it be addressed to *individuals* themselves by extending their opportunity sets (Sugden 2004, 2018), or is there perhaps another alternative? In Lecouteux and Mitrouchev (2023) we propose an alternative that I here briefly present.

In order to better grasp the different standpoints to which BPP can be addressed, let us replace paternalism with "the view from nowhere", anti-paternalism with the "view from somewhere", and explain what is meant by that. The "view from nowhere" is a term of Sugden (2018) borrowed from Nagel (1986) to identify paternalism as a policy approach, which takes the standpoint of a *social planner* (or a practical policymaker in BPP). Because individuals are not the actors of their own welfare evaluation, Sugden argues that it is a situation to be seen from *nowhere* in the society that is being assessed. His approach, in contrast, is a "view from *somewhere*", because when individuals engage in mutually beneficial transactions, they are actors of what goods to exchange, in which quantities, at which price, in which conditions, etc. In this sense, it is a "view from somewhere" in the society that is being assessed through the level of mutual benefit it produces (to be measured by e.g. the number of voluntary transactions in a market economy).

Like Sugden, we recognise that for many cases, the role of an external third party to evaluate welfare (the

[6] There is a deep debate in metaethics between *cognitivism*, which argues that ethical beliefs can be "true" or "false" propositions, and *non-cognitivism*, which argues that they cannot. For obvious reasons, I bracket out this debate and refer the reader to Van Roojen (2023).

social planner or policymaker) is questionable.[7] However, the problem we see with Sugden's approach is that, if normative authority is fully granted to individuals, any psychological features that may play out a significant role in their welfare is put aside. But one may not always have (for example) full information before making a choice (like engaging in a voluntary transaction). As a result, one may be disappointed of the outcome of such a transaction. That is, it may not result in a *mutually beneficial* exchange. Sugden's normative theory is valid as long as individuals perfectly know what makes them better off (which does not mean they are fully rational). We share doubts about this when we think of the many cases where preferences are either incomplete, or (as we consider it in our framework) *context-dependent*.

By context dependency, we mean a situation where *any* kind of properties may affect one's choice, e.g. one's mood of the day, the weather, the location of the alternatives, etc. In order to avoid paternalism, which is a situation where the social planner needs to identify (perhaps arbitrarily) what contexts and preferences are normatively "problematic", we argue that the role of the external observer should not be to give to individuals what they would choose if they were "enlightened". Instead, we argue that if such a social planner exists (in the role of a policymaker in BPP), he/she needs to ensure that social institutions are designed in a way that individuals have the opportunity to form their own enlightened judgements about what is best for themselves. With this approach, we aim to escape the problems of both paternalism (view from nowhere) and non-paternalism (view from somewhere), and then to think about the policy tools aligned with this normative theory (*boosts*).

How to illustrate this third alternative in practice? Imagine the following scenario, taken from Lecouteux and Mitrouchev (2023). You and your colleagues regularly go to the same restaurant for lunch. You know there are often very tasty desserts, although they are a bit oversized for you. You therefore face a recurrent choice problem. Should you take a dessert, knowing that it will give you a significant and immediate enjoyment but probably make you a bit sleepy in the afternoon? Your actual choice is very likely to depend on the context: whether some of your colleagues have already taken a dessert, whether you are in a good or bad mood, whether the location of the desserts are nicely displayed on the counter to arouse your appetite, etc. Since your preference for a dessert is *context-dependent*, it is not clear whether your actual choice constitutes a reliable evidence of what maximises your welfare – and therefore, whether it is the "good" choice in such a situation. We can distinguish between three different approaches to define what your own judgement is.

(i) It is your current judgement in a given choice situation, e.g. it is a rainy day, you have just got a paper rejected, and the dessert looks very nice on the counter, so you take it.

(ii) It is the outcome of a kind of bargaining between your current self and what you imagine your future self will be this afternoon, weighing the pros and cons of both options, and whose result will very likely be context-dependent too.

(iii) It is your counterfactual "enlightened" judgement about the choice problem, i.e. the rational trade-off between an immediate enjoyment and a later cost, the result of which being determined by your deeply held values and preferences for different personal aspects.

While the "view from somewhere" (anti-paternalism) would hold that judgement (i) is best, the "view from nowhere" (paternalism) would hold that judgement (iii) is best. In contrast, our proposition is that judgement (ii) is best. The originality of our approach is that we give importance to the individual's *ability* to confront different judgements he/she may have on what is constitutive of his/her welfare. In this regard, our proposition is *cognitive*. We acknowledge there is inherently no context-independent "best alternative" because every choice is context-dependent – as Thaler and Sunstein (2003) rightly emphasise it.

[7]This depends on the emergency of the problem, to which public policy is concerned about. The obvious case is ecology, in which case paternalism, and even coercion, may be justified.

In our normative theory, individuals need to have a minimal ability of self-reflection, which gives them the possibility to imagine themselves in another context, and then to form counterfactual judgements about how themselves, in different contexts, would choose. According to the "view from *manywhere*", not only each of your "self" matters, but each of his/her ability to imagine the choice problem from different perspectives. This eventually gives a normative authority to yourself across time. We call our normative approach the "view from *manywhere*" because it evaluates welfare from the standpoint of *many* perspectives of an individual (yet not *all* perspectives, as this is practically impossible).

The "view from *manywhere*", however, poses some challenges. In particular, one should not forget the *practical* goal of policy. *Nudges* have been extremely successful in BPP because they are (i) very simple to understand, (ii) efficient (i.e. they lead to a change of the desired behaviour) and (iii) they can be implemented at a very low cost (e.g. drawing flies in airports' urinals). Note also that (at least in the contemporary societies we live in) real-life policies *are* about policymakers (either elected or assigned) who choose for society. While our approach does give a role to the policymaker in designing contexts where individuals would be better able to confront different perspectives, it is less clear how to do so in real-life policies. As we put it in Lecouteux and Mitrouchev (2023), we see the application of our theoretical framework in the *boost* agenda (Grüne-Yanoff and Hertwig 2016). Contrary to *nudging*, whose aim is to *exploit* cognitive biases to influence behaviour, *boosting* aims to expand individuals' abilities to make choices that are aligned with their goals. Leaving these practical concerns apart, I now move on to another philosophical challenge of BPP, which is about the direct *addressee* of BPP: the *individual* himself/herself.

3 How to Evaluate Welfare when Choices Are Inconsistent over Time?

3.1 The Identity Problem of Individuals

Let us start with a provoking question. If (by definition of the standard framework) an economic agent is an individual represented by a set of *consistent choices* over alternatives but (from what we know of empirical research) his/her choices are inconsistent, can we say that an "economic agent" still exists? The answer appears obvious: we simply have to drop the consistency principle to rescue the concept of economic agent. That is, the *behavioural* definition of an economic agent would simply be an individual represented by a set of choices over alternatives *that are not necessarily consistent*. So far so good, at least for descriptive purpose. Indeed, it should not bother one to observe that choices are inconsistent when studying a given phenomenon in an experimental setting, and then to model such a behaviour.[8]

But how about if we have to evaluate welfare from inconsistent choices, i.e. when the purpose is not *descriptive* but *normative*? In other words, if choices are inconsistent, how to know what makes people better off? While the previous question was about the "right" *standpoint* of BPP, this question is about the *identity* of individuals, who are directly concerned about BPP. Note that providing an answer to this second question does not depend on how we answered the first. Whether one holds that BPP should be addressed to (i) the social planner/policymaker (view from nowhere), (ii) the individual himself (view from somewhere), or (iii) the individual accounting for many perspectives (view from *manywhere*), the individuals whose welfare is being assessed have to remain *the same* individuals for their welfare to be assessed. Thus, the identity question particularly arises when choices are inconsistent *over time*. This problem was well emphasised by Kahneman (1994, 31): "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

[8] In that vein, prospect theory was originally proposed by Kahneman and Tversky (1979) as an alternative *descriptive* model of individual behaviour to expected utility theory, as well as hyperbolic (or quasi hyberbolic) discounting models (Laibson 1997) as alternative *descriptive* models of intertemporal choice.

experienced in the future?".[9] Thus, if we aim to attribute an identity to the economic agent, it should not be based on a consistency principle (such as *completeness*, *transitivity* or *invariance*) but on something else. The question is, on what?

The approach we take in Mitrouchev and Buonomo (2023) is *ontological*. We argue that the *ethical* question of how welfare should be evaluated over time is *conditioned* by the answer provided to the *ontological* question of what grounds identity over time. The reason is simple. If we cannot state that an individual remains the same from one time to another, evaluating his/her own welfare would be as if evaluating the welfare of two different individuals. That is, in order to evaluate the welfare of individuals, we first need to be confident that those individuals actually remain the same over time.[10] Depending on how we answered the ontological question, it will impact the way we evaluate one's welfare. For example, assume we find out that individuals persist over time in virtue of their *psychological* relations such as preferences. If that is the case, because their identity would be constituted by their *preferences*, it would seem natural to evaluate their welfare based on their *preferences* (and not something else that is external to their psychology).

In order to investigate this ontological inquiry, let us start by defining a general criterion of identity over time, as it is defined in the literature of personal persistence. 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 $\Phi(x, y)$, where = is the relation of numerical identity over time, and Φ is the constitutive condition whereby the identity of x and y is determined.[11]

In the literature of personal persistence, there are different theories on what Φ is. A theory of personal persistence either grounds identity on something (and the ontological inquiry is specifically to find out what this "something" is) or nothing (thus concluding that persistence over time, and therefore identity, does not exist). The *psychological* theory states that an individual is identical over time by virtue of some *psychological aspects* such as memories, intentions, beliefs, goals, desires and similarity of character. The *physical* theory states that an individual is identical over time by virtue of some *physical aspects*, such as body parts (typically, physical parts in the brain). The *narrative* view is an extension of the psychological view, stating that an individual persists over time in virtue of some *psychological relations that can be collected together* in a "meaningful" story. Last but not least (what we refer to as) the *sociological* view states that *social factors* that characterise an individual born in a given environment (his/her culture, norms, habits) form his/her ontological unit that gradually becomes responsible for and concerned with its own future.[12]

What are the possible implications of these theories of personal persistence for welfare evaluation? Assume the narrative theory is correct, that is, that x and y are the same person over time because they are connected by some self-told narrative relations. In this case, it seems that any ethical rule based on such a narrative structure would make sense, because it would refer to something that is ontologically grounded. This "something" would be 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. In this sense, the narrative theory of personal persistence could support our proposition in Lecouteux and Mitrouchev (2023), where the individual has an ability to self-reflect the experience of choosing a different alternative in a different context. Assume now the sociological theory is correct, according to which x and y are the same

[9]See also Schelling (1980, 94-118), Braga and Starmer (2005, 60), Sugden (2010, 54) and Sunstein (2019, 69-75), who point out this important problem.

[10]This question may appear too philosophical for the practice of BPP, especially if the time interval which concerns the policy is relatively short. Also, individuals are being asked what they prefer at t_1 for a given policy, but their delayed preferences (e.g. at t_2, t_3, ...) are usually not taken into account if they change after the implementation of the policy at t_2, even if those delayed preferences should be (for whatever reason) granted more normative authority. In our theoretical framework, we assume it can be possible to account for one's preference at any point in time.

[11]We add *informative* conditions of this criterion of identity, which are (i) *non-triviality*: Φ should have a different meaning from, or at least is not logically equivalent to the identity it constitutes, (ii) it should be logically possible that x and y do not satisfy Φ, and (iii) Φ should not presuppose the identity it should demonstrate.

[12]See Shoemaker (2019) for a survey on the relationship between the main theories of personal persistence and ethics.

person over time because they are connected by some sociological relations such as norms, habits, etc. In this case, it seems that any ethical rule based on the characterisation of some institutions (such as those that promote the free market) would have a significant theoretical advantage over other rules. This is because those ethical rules could be explained at the ontological level, i.e. in this case, that personal persistence is a matter of sociological features such as norms and habits.

However, in Mitrouchev and Buonomo (2023) we do not provide any argument on whether one theory of personal persistence is superior to another.[13] This is in fact a complex debate in the literature of personal persistence, where more needs to be investigated in order to provide convincing arguments for justifying one normative approach (e.g. paternalism or anti-paternalism) over another.

3.2 Going "Ontological" All The Way Down

Still, more can be investigated about the ontology of individuals for normative purposes (here BPP, but not exclusively). To our knowledge, all the theories of personal persistence that are implicitly assumed in the literature concerned with the normative implications of behavioural economics (whether they consider the individual as a collection of "multiple selves" or a "unified self" over time) represent *time* as an exogenous variable of the persisting individual. In the identity problem formulated by Kahneman (1994), I is composed of temporal selves who are assumed to be parts of I. One difficulty, which to our knowledge has not been tackled in the literature of economics-and-identity, is the relationship between temporal selves and temporal parts. In fact, formulating the problem as Kahneman (1994) already makes a presumption about identity – that is, temporal selves are somehow coextensive to temporal parts. But instead of focusing on whether the individual can either be represented by multiple selves or a unified self over time, we believe there is more interest in focusing on the relationship between *parts of persons* and *time*.

The literature which encapsulates this debate roughly divides in two competing theories: *endurantism* and *perdurantism*. According to endurantism, physical entities persist over time passing through time and being, strictly speaking, identical over time. That is, to say that I persists over time by enduring means that given two different times t_i and t_j, I at t_i and I at t_j is the same *entire* (or numerically identical) entity respectively at these two different times. By contrast, perdurantism states that physical entities persist by having different temporal parts at different times. Just like our common sense idea that concrete entities are composed of different spatial parts located at different regions of space, perdurantism claims that they are also composed of different temporal parts located at different regions of time. Thus, according to perdurantism, concrete entities (among which living entities like individuals) are not only extended in space but also in time.

Consider the following example of endurantism. When we claim that "I was in Paris at a conference three days ago", it was I who saw her colleagues three days ago and who was happy to present her research. Today, I is at home. When she took the train on her way back to her home, she similarly crossed time. The point is that it is not just a part of I that is at home today, with memories of her conference in Paris. Instead, it is the *whole* I, i.e. the same individual who was at the conference three days ago. The endurantist account of persistence sounds rather intuitive. Indeed, it is actually well in line with the way we ordinarily think about ourselves in the world. In this matter, it may support the unified view of the self. But consider now an example of perdurantism. I, from the time she arrived in Paris to the time she came back home is composed of several spatial parts, such as her head, arms, legs, and so on. What perdurantism argues is that the four-dimensional individual I has temporal parts as well, such as I-on-Monday, I-on-Tuesday and I-on-Wednesday. This means that a temporal part of the four-dimensional individual I is I during an interval of time which is included in I's temporal boundaries, namely between her departure to Paris

[13] In our analysis, we categorise Davis' (2011) theory of identity as a *sociological* theory of personal persistence – to which we are most sympathetic with.

to her arrival back at home. More generally, if a spatial part of an individual I is a part of I which is smaller than I in some spatial dimension(s), a temporal part of I is a part of I that is shorter along the temporal dimension, but which, during the relevant temporal interval, has the same spatial extent as I – i.e. it overlaps everything that is part of I during the relevant temporal interval.

A useful point to be added in a further study would be to put into question the presupposition of how individuals persist over time. In particular, what are the implications of a "temporal part" representation of identity in economics? If perdurantism is true, how should we treat temporal parts of individuals who make economic choices? As Davis (2011) puts it, it is a fact that intertemporal choice is largely studied in economics from both descriptive and normative aspects. But it may appear disputable to care about one's intertemporal choice (like in BPP) if one is not the same temporal part of individual from one period to another. So what if I is composed of several parts that are extended through time, but that she is not, strictly speaking, the whole I at each slice of time?

The point is if perdurantism is considered to be the "right" theory of identity, it may provide an ontological defence for the multiple selves view endorsed in e.g. libertarian paternalism. Another point of Davis (2011) is that life plans such as education, investment or health involve choices over extended selves that seem related to each other. The ontological debate between endurantism and perdurantism may then enlighten our understanding of how selves actually relate to each other. Lastly, Davis (2011) underlines the point that individuals have a capacity to make a choice in time, which means there is potentially "something enduring about them apart from all their choices" (6). Because we largely agree with Davis (2011) about these three points but not necessarily whether there is something *enduring* about individuals, the philosophical debate about identity in economics has the merit of being established in a more promising framework.

4 Concluding Remarks

In this chapter, I aimed to emphasise two philosophical challenges of BPP. The first is: *to whom should BPP be addressed?* The second is: *how to evaluate welfare when choices are inconsistent over time?* Both questions involve important philosophical reasoning on what *policy* and *identity* are (ultimately) about.

But there is more. Reading this chapter, the careful reader may have wondered why social choice theory, which is the discipline that is specifically devoted to analyse how preferences can be aggregated – and pointing out all associated problems with different possible aggregation rules – did not play a role in my analysis. If preferences not only change across *different* individuals (interpersonal relations) but also across the *same* individuals (intrapersonal relations), the tools of social choice theory should also apply. This is absolutely right, knowing that there is currently no study (of my knowledge) aiming to fill the gap between BPP and social choice theory. In what follows, I suggest some perspectives of further research to the interested reader who may want to investigate these issues.

The possibility to make interpersonal comparisons of utilities and the compatibility of values, are, so far, absent from debates about the philosophical challenges of BPP.[14] As a consequence, questions such as what informational basis and aggregation rules should be considered in order to evaluate social welfare constitute a gap in the literature.[15] These questions are specifically salient when policymakers hold a restricted set of values for welfare evaluation. Consider libertarian and asymmetric paternalisms. In the

[14]Regarding interpersonal comparisons of utilities, an exception is Kahneman (1999), who made a substantial effort to provide a large amount of psychological evidence about the possibility to make interpersonal comparisons of utilities. Regarding the compatibility of values, Sugden (2004, 2018) only accounts for freedom to choose as the informational basis in his normative theory. Therefore, he escapes from the potential conflicting relationship between freedom and welfare. This problem, however, applies to libertarian paternalism, which accounts for freedom *and* welfare. Consequently, it also applies to BPP.

[15]See also Baujard (2015, 303), who points out the problem of libertarian paternalism in not being explicit about what aggregation rule should prevail. Note that Thaler and Sustein (2003, 178) briefly mention a majority criterion, but do not develop it further.

line of the welfare tradition, these approaches only take Pareto efficiency as the welfare criterion. There are however other values such as liberty, autonomy, freedom, equality, fairness, etc., that are worth being taken into account. The point is that if we only account for Pareto efficiency, we ignore all the other values that matter to individuals – but this a known problem emphasised by Sen (1970).

In order to build bridges between BPP and social choice theory, an interesting study would be to check if Sen's impossibility of a Paretian Liberal also applies at the individual level. Can we combine the Pareto criterion with respect to individuals' liberty to choose whatever they want, knowing that individuals make inconsistent choices? If we assume an individual to be a collection of subpersonal selves with a set of preferences or strategies, it is in fact very likely that any kind of paradox or impossibility theorem known in social choice theory would also apply at the individual level.[16] In this regard, impossibility theorems and related paradoxes of social choice, such as the famous theoretical results of Arrow (1951) and Sen (1970), could, in a way, also apply to BPP.[17]

I hope to have provided the reader with some thoughts on how philosophical reasoning can help us advancing on some challenges in BPP. I also hope I have stimulated him/her with some valuable directions of research.

References

Arrow, Kenneth J. 1951. *Social Choice and Individual Values*. First ed. John Wiley & Sons.

Baujard, Antoinette. 2015. "Beyond the Consent Dilemma in Libertarian Paternalism, a Normative Void." *Homo Oeconomicus* 32 (2): 301–5.

Bernheim, B. Douglas, and Antonio Rangel. 2007. "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics." *American Economic Review* 97 (2): 464–70.

———. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal of Economics* 124 (1): 51–104.

Binder, Constanze. 2014. "Plural Identities and Preference Formation." *Social Choice and Welfare* 42 (4): 959–76.

Buchanan, James M. 1964. "What Should Economists Do?" *Southern Economic Journal* 30 (3): 213–22.

———. 1987. "The Constitution of Economic Policy." *The American Economic Review* 77 (3): 243–50.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"." *University of Pennsylvania Law Review* 151 (3): 1211–54.

Coase, R. H. 1960. "The Problem of Social Cost." *Journal of Law and Economics* 3: 1–44.

Davis, John Bryan. 2011. *Individuals and Identity in Economics*. Cambridge University Press.

Gilboa, Itzhak, and David Schmeidler. 2001. *A Theory of Case-Based Decisions*. Cambridge University Press.

Grüne-Yanoff, Till. 2012. "Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles." *Social Choice and Welfare* 38 (4): 635–45.

Grüne-Yanoff, Till, and Ralph Hertwig. 2016. "Nudge Versus Boost: How Coherent Are Policy and Theory?" *Minds and Machines* 26 (1-2): 149–83.

Halpern, David. 2015. *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. Allen.

Hédoin, Cyril. 2015. "From Utilitarianism to Paternalism: When Behavioral Economics Meets Moral Philosophy." *Revue de Philosophie Économique* 16 (2): 73–106.

[16] See Binder (2014), who characterises the conditions under which the aggregation is possible (or not) at the intrapersonal level.

[17] Consider also the Coase (1960) theorem. As underlined by Hédoin (2015, 78), behavioural economists have not considered potential results, where an individual's multiple selves bargain over his internalities (i.e. externalities at the intrapersonal level). A contribution of the Coasean solution applied to libertarian paternalism may be of good use, simply to show that the social cost problem also applies at the intrapersonal level. That is, if internalities are nothing more than externalities at the intrapersonal level, one may not find it surprising that the conclusions of *social* bargaining also apply to *individual* bargaining.

Kahneman, Daniel. 1994. "New Challenges to the Rationality Assumption." *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift Für Die Gesamte Staatswissenschaft* 150 (1): 18–36.

———. 1999. "Objective Happiness." In *Well-Being: The Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener, and Norbert Schwarz, 3–25. Russell Sage Foundation.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–92.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *The Quarterly Journal of Economics* 112 (2): 375–406.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics* 112 (2): 443–78.

Lecouteux, Guilhem, and Ivan Mitrouchev. 2023. "The View from Manywhere: Normative Economics with Context-Dependent Preferences." *Economics & Philosophy*.

Mill, J. S. 1859. *On Liberty*. John W. Parker & Son: London.

Mitrouchev, Ivan. 2024. "Normative and Behavioural Economics: A Historical and Methodological Review." *The European Journal of the History of Economic Thought*.

Mitrouchev, Ivan, and Valerio Buonomo. 2023. "Identity, Ethics and Behavioural Welfare Economics." *Economics & Philosophy*.

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford University Press.

Oliver, Adam. 2013. *Behavioural Public Policy*. Cambridge: Cambridge University Press.

Scoccia, Danny. 2019. "Paternalisms and Nudges." *Economics and Philosophy* 35 (1): 79–102.

Sen, Amartya. 1970. "The Impossibility of a Paretian Liberal." *Journal of Political Economy* 78 (1): 152–57.

———. 1979. "Utilitarianism and Welfarism." *The Journal of Philosophy* 76 (9): 463–89.

Shoemaker, David. 2019. "Personal Identity and Ethics." *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/identity-ethics/.

Smith, A. 1759. *The Theory of Moral Sentiments*. A. Millar: London.

Sugden, Robert. 2004. "The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences." *American Economic Review* 94 (4): 1014–33.

———. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press.

Thaler, Richard H., and Cass R Sunstein. 2003. "Libertarian Paternalism." *American Economic Review* 93 (2): 175–79.

Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Revised and expanded ed. Penguin Books.

Todd, Peter M., and Gerd Gigerenzer. 2012. *Ecological Rationality: Intelligence in the World*. Oxford University Press.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (2): 207–32.

———. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.

Van Roojen, Mark. 2023. "Moral Cognitivism Vs. Non-Cognitivism." *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/moral-cognitivism/.

Wakker, Peter. 2010. *Prospect Theory for Risk and Ambiguity*. Cambridge University Press, Cambridge.