Libertarian Paternalism and Susan Hurley's Political Philosophy

Ittay Nissan-Rozen[1]

Philosophy/PPE

The Hebrew University of Jerusalem

Ittay.nissan@mail.huji.ac.il

[1] I thank Hadrian Mamou and an anonymous referee for their helpful comments on previous versions of this paper. My research was supported by the Israeli Science Foundation, Grant Number 327/18.

Abstract

As the use of nudges by governmental agencies becomes more common, the need for normative guidelines regarding the processes by which decisions about the implementation of specific nudges are taken becomes more acute. In order to find a justified set of such guidelines one must meet several theoretical challenges to Libertarian Paternalism that arise at the foundational level. In this paper, I identify three central challenges to Libertarian Paternalism, and suggest that Susan Hurley's political philosophy as presented in her Natural Reasons *(1989) can be viewed as offering powerful responses to them.*

Key Words

Libertarian Paternalism; Nudge; Susan Hurley; Personal Autonomy

0. Introduction

Since the introduction of the term "Libertarian Paternalism" (by Thaler and Sunstein in a series of papers, including 2003 and 2008), both the view to which the term refers and the set of policy tools usually associated with it have drawn several criticisms. Attempts to accommodate these criticisms (by Thaler and Sunstein, and by others[2]), which are based on differing understandings of key theoretical terms and involve differing theoretical commitments, make it clear that Libertarian Paternalism (LP) is better understood not as a single view, but rather as a family of possible (and usually implicit) political views that share some common features.

Following Thaler and Sunstein 2008, I take any view that is committed to the claims that at least in some cases it is possible for governments to influence the behavior of their citizens without compromising the citizens' freedom, and that in some of these cases governments ought to do so, to belong to this family. This characterization leaves plenty of room for different and even incompatible political views to be properly described as "Libertarian Paternalistic views". Different Libertarian Paternalistic views can differ in the scope of cases in which they allow for paternalistic interventions, in the justifications they give to such interventions and in the interpretations they give to several ethically-laden concepts often used in these justifications.

In this essay, I argue that the conception of liberal democracy presented by Susan Hurley (1989), a conception that she describes as one "which gives the value of autonomy the distinctive role Dworkin gives to the value of equality" (Hurley 1989, p.

[2] See for example Bovens 2009, Hausman and Welch 2010, Wilkinson 2013, Saghai 2013a and 2013b, Mills 2013 and 2015 and Lepenies and Maleca 2015, Sunstein 2016, Engelen 2017, Schmidt 2017 and 2020.

324), can be used as an especially attractive normative platform for LP. Hurley's view of liberal democracy should be attractive to libertarian paternalists, not only because of the powerful responses to several widely-held criticisms against LP (and the set of policies usually associated with it) it can be viewed as suggesting, but also because these responses are unified in the sense of emerging from a single coherent picture of the social world.

In section 1, I argue that Hurley's political view manages to do this because – much like Thaler and Sunstein's original characterization of LP – it emerges from a specific view regarding practical rationality and the role it plays in explaining human behavior. By so arguing, I hope to demonstrate how relevant Hurley's philosophy – which despite being highly praised has not yet managed, I believe, to achieve the influence it merits – can be to current debates in political philosophy.

When presenting his own responses to some of the criticisms that will be discussed here, Sunstein writes:

> To come to terms with the ethical questions, it is exceedingly important to bring first principles in contact with concrete practices. For purposes of orientation, it will be useful to give a more detailed accounting of potential nudges that might alter choice architecture. One reason is to avoid the trap of abstraction, which can create serious confusion when we are thinking about regulatory (or other) policy. As we shall see, the ethical evaluation of nudges depends on their concrete content, not on their status as nudges. (Sunstein 2015b, p. 242)

While I agree with Sunstein that such a methodological approach is appropriate when dealing with the justification of the use of nudges, in this essay my task is different. My task here is to offer firm philosophical foundations to LP that will enable one to draw a different type of policy recommendations. The policy recommendations to which accepting my suggestion leads are much more general in nature. They are mainly concerned not with accepting specific (types of) interventions, but rather with the appropriate institutional design in which decisions about which interventions are adopted and how exactly they are implemented are taken.

As the use of nudges by governmental agencies becomes more common, the need for normative guidelines regarding high-order ethical decisions having to do with nudges (i.e. decisions in which the pressing normative questions do not concern the justifiability of specific nudges but rather concern the justifiability of the process through which decisions about potential nudges are taken) becomes more acute. However, in the literature most of the normative discussion concerning nudges deals with first order ethical questions.[3] In order to answer high-order ethical questions regarding nudges, one must address questions regarding the ideological foundations of LP (even if, as I discuss in section 4, LP should not be viewed as a comprehensive political ideology on par with social democracy or libertarianism). In this essay I use Hurley's philosophy to take a first step in this direction.

The rest of the essay is organized as follows. In section 1 I present three central challenges that any plausible account of LP must handle: a conceptual challenge, an ethical challenge and a political challenge, and explain how Hurley's philosophy is able to meet these three challenges in a unified way. In sections 2-4 I discuss each one of these challenges in more depth.

[3] Most, but not all. Schmidt 2017 is one notable exception. Schmidt’s conclusions regarding policy recommendations are very close to the ones I draw here. However, the justificatory basis I use here and the conception of liberal democracy on which it is based are very different from the ones he is committed to.

1. Three challenges for Libertarian Paternalism

How is LP possible? Thaler and Sunstein’s (T&S) famous idea is that the fact that real people are not perfect rational agents (what they call "Econs") opens the door for paternalistic interventions that do not conflict with the freedom of choice.

Their idea is by now well known. Choices are always made in a specific environment and presented to the chooser in a specific way. T&S call this "the choice architecture". Now, while the choices of perfect rational agents would never be affected by the choice architecture in which they are made (I discuss this claim in more depth below), different choice architectures do have different effects on the way *real people* choose. Thus, governments (or other bodies) can shape choice architectures to influence people's behaviour without limiting their freedom to choose any of the alternatives available to them.

T&S call such interventions "nudges". They present two different definitions of the term and argue that the two definitions have the same extension. The first definition takes a nudge to be: "any aspect of the choice architecture that alters people’s behaviour in a predictable way without forbidding any option or significantly changing their economic incentives" (Thaler and Sunstein 2008, p. 6). The second definition takes a nudge to be: "any factor that significantly alters the behavior of Humans, even though it would be ignored by Econs" (Thaler and Sunstein 2008, p. 9). LP can thus be viewed as the position according to which governments should use nudges in order to promote their citizens' welfare, *as judged by the citizens themselves*.

Many people find LP's central demand to be highly intuitive and many specific nudges that have been implemented in different public systems or merely suggested in the literature enjoy great public support (for evidence see Sunstein 2016). However, it has also attracted a significant amount of criticism. In this essay I do not attempt to defend LP from all the criticisms directed at it (see Barton and Grune-Yanoff 2015 as well as Schmidt and Engelen 2020 for good reviews, and see Sunstein 2015a, 2015b and 2016 for a list of replies). Rather, my aim here is to ground (at least one possible version of) LP on firm philosophical foundations. In order to do so it will be instructive to concentrate on three central challenges that any plausible version of LP should be able to meet. The first challenge is conceptual, the second, ethical and the third, political.

In the next sections I present in detail each one of these three challenges and the way they can be met by adopting the Hurleyian picture. It will be instructive, however, to first provide the reader with a rough "bird's eye" view of Hurley's theory and the way in which it can be viewed as responding to the three challenges which are as follows.

The conceptual challenge is that of drawing a line between a choice option and the environment in which the option is located. Without the possibility of drawing such a line, it is not clear whether any given specific intervention counts as a nudge. Is the location, Y, of a food product, X, on the shelves in a store part of the description of the choice option "X located in Y" or is it part of the environment in which the choice option "X" is located? Giving different answers to this question can lead to committing oneself to different judgements regarding whether regulating the location of unhealthy food products in grocery stores counts as a nudge.

There are several ethical challenges discussed in the literature with respect to LP. Here, however, I will concentrate on what I take to be the most pressing one, which is the threat the use of nudges poses to personal autonomy. This threat, arguably, arises because when a specific nudge succeeds in effecting a person’s decision, it seems that this decision is not the result of rational deliberation and thus not fully autonomous.

The political challenge is that of giving an account of LP that goes beyond a commitment to using nudges as local solutions to local problems. Even if LP should not be understood as a comprehensive political ideology on par with libertarianism and social democracy, it is grounded in (and committed to) a specific set of values and a specific picture of the social world that cohere in a way characteristic of political ideologies. Any attempt to answer higher-order ethical questions having to do with the use of nudges should be based on an understanding of the nature of this ideological content.

How can these three challenges be met using Hurley’s philosophy?

In *Natural Reasons* (1989), Hurley began her exploration in the domain of philosophy of mind and ended it in the domain of political philosophy. She wrote:

> A major thesis of the book is that arguments drawn from the philosophy of mind may be used to undermine widely-held subjectivist positions in ethics and associated positions in politico-economic theory. (p. 3)

The conclusions Hurley arrived at on the political level are not only inspired by the relevant literature in the philosophy of mind, but are also designed specifically to solve conceptual problems that arise when philosophers of mind think about the notion of rational agency. Thus, the relevance of Hurley's philosophy to LP is certainly not accidental: LP and Hurley's philosophy have the same starting point (the conception of individual rationality/irrationality), and they are developed in the same direction (that of political philosophy).

However, while the starting point of T&S is a reliance upon an implicit assumption that the concept of rationality used by social scientists who investigate human decision-making processes is relatively well-defined, and so they do not seek to analyse it, Hurley made no such assumption. On the contrary, she explicitly explored the conceptual commitments the notion of rationality as used in the social sciences carries with it (this may be partly due to the fact that at the time Hurley wrote her book the relevant psychological literature was only at the beginning of its second decade and conceptual issues were thus more pressing). In other words, Hurley directed her efforts to constructing exactly what (so I will argue) is missing from contemporary notions of LP, namely firm conceptual foundations.

It is, thus, unsurprising that Hurley's conception of liberal democracy can potentially offer powerful responses to the three challenges outlined above. Before delving into the details, I will outline rough sketch of how I believe Hurley's philosophy can be so interpreted.

The conceptual challenge is met by presenting an account of how to draw the line between a choice-option and its environment; of what, in other words, makes an individuation of the alternatives in a given choice situation an admissible one. According to Hurley's account, admissible individuations, for an agent, are the ones dictated by the most coherent theory that displays the agent as a rational agent, who responds to reasons in an appropriate way. Such a theory must be sensitive to the features of reality that the agent herself takes to be reasons for action, but at the same time, the agent might make mistakes when trying to figure out how this theory looks (e.g. when trying to figure out which individuations are admissible)[4].

[4] As will be explained, this feature of Hurley's account, i.e. the possibility of one being mistaken about which features of reality one values, makes it especially fit for the purpose at hand. It gives Hurley's

Hurley takes the process of deliberation in which one looks for the most coherent theory of oneself as a rational agent to be the process through which one exercises one's personal autonomy. Personal autonomy, according to Hurley, is not about choosing in accordance with some privileged set of attitudes. Rather, it is about constituting oneself as a rational agent through a process of self-interpretation. Thus, although irrational behaviour is systematically related to lack of personal autonomy, it is not the case that just any instance of irrational behaviour poses a threat to one's autonomy. On the contrary, it is the possibility of irrationality, which triggers this process of deliberation, that enables one to exercise one's autonomy. Thus, successful nudges in themselves do not necessarily pose a threat to personal autonomy (as understood by Hurley) as long as they are implemented in a way that enhances people's tendency to get engaged in such a process of deliberation. Hurley's conception of personal autonomy opens the door, then, to implementations of nudges that do not threat people's autonomy.

According to Hurley, the process of deliberation in which an agent looks for the most coherent theory of herself as a rational agent, the process through which one exercises one's autonomy, must be (for reasons that will be discussed) social in nature. One can be autonomous, one can make rational choices, only by relying on others. Thus, if autonomy is an important value that should be protected, the political system has to be arranged in a way that allows society to play an adequate role in the citizens' process of deliberation. Hurley's conception of liberal democracy is of a democratic system that respects this restriction. In this sense it meets the third challenge introduced above, the political one.

account of personal autonomy a flexibility that another important and highly influential cohertanist account of personal autonomy, that of Ekstrom (1993), lacks (although the two accounts are close in many respects). Space limitations do not allow me to discuss Ekstrom's account (and other important accounts of personal autonomy) in detail here.

I now turn to a more detailed discussion of both the three challenges and the way Hurley’s philosophy can be used to meet them.

2. The conceptual challenge

a. Presenting the conceptual challenge

A good way to present the conceptual challenge is by explaining why the two definitions T&S suggest for "nudge" have the same extension. Notice, first, that in order for the two definitions to be equivalent it must be the case that a feature of the environment that belongs to the choice architecture can have no causal effect on the way a perfectly rational agent chooses. A feature of the environment that can affect the way a perfectly rational agent chooses should instead be treated as part of the description of the alternatives in the choice set: it should, in other words, be described as *something a rational agent might care about.*

For example, in order to argue that fixing the default choice for couples about to get married as "sign a prenuptial agreement" is an intervention that should count as a nudge (as suggested by T&S in Thaler and Sunstein 2008, chapter 3), T&S's first definition demands that signing a prenuptial agreement *when this is the default option* is in fact *the same alternative* as signing a prenuptial agreement when the default option is the state law. Otherwise, by changing the default option, the government necessarily eliminates from the choice set one alternative (e.g. signing the agreement when this is the default option) and replaces it with another (e.g. signing the agreement when the default option is the state law).

T&S's second definition demands that a perfectly rational agent would choose the same option regardless of whether signing the prenuptial agreement is the default option or not. Thus, if the two definitions are equivalent, it must be the case that if a perfectly rational agent could[5] choose differently in the case in which signing the agreement is the default option than they would choose in the case it is not, then signing the agreement when the agreement is a default option is a different alternative from signing the agreement when signing it is not the default option.

T&S's second definition, it is easy to see now, serves as a criterion for their first definition. The phrases "forbidding an option" and "significantly changing their economic incentives" are to be understood in the following way: an intervention forbids an option or changes an economic incentive iff it *can* have an effect on the behaviour of a perfectly rational agent. However, the question arises: what is the criterion for distinguishing interventions that can have an effect on the way a perfectly rational agent chooses from interventions that cannot? In other words, what is the criterion for deciding whether a specific feature of reality belongs to the choice architecture or to the description of the alternatives in the choice set?

This problem – which arises not only in the context of nudges – is well known to decision theorists[6] and, as we will see, serves as one of Hurley's starting points in developing her account of liberal democracy. The problem is, in fact, quite general. Practical rationality (as usually understood in the literature) does not put any restrictions on single attitudes. It only restricts agents' attitudes (preferences, beliefs and desires)

[5] But not necessarily "would". This is an important difference. A rational agent does not *have* to care about whether or not signing the prenuptial agreement is the default option, and she certainly does not have to care about that so much that whether or not it is the default option changes the way she chooses. However, it should be *rationally permitted* for her to do so.

[6] For discussions other than Hurley's, see Broome (1991, 1999), Bhattacharyya, Pattanaik and Xu (2011), and Fumagalli (2020).

directed to certain acts or outcomes given other attitudes those agents have that are directed to the same acts or outcomes.

For example, transitivity of preferences is usually taken to be a condition of practical rationality. It demands that if an agent prefers one alternative, A, to another alternative, B, and also prefers B to a third alternative, C, she must prefer A to C. This condition put a substantial restriction on the agent’s preferences only by virtue of the possibility of the same alternative (in our case, alternative A) being considered in two different contexts (once when compared to B and once when compared to C).

The possibility of considering the same alternative in two different contexts arises only if rationality forbids agents from caring about certain features of reality. Otherwise, there can be no violations of transitivity, as strictly speaking, no two alternatives are completely identical. Alternatives are always presented to agents at different points in space and time, out of different menus and in different modes of presentation. The same point holds with respect to the other standard conditions of practical rationality (see for example Broome's 1991 and 1999 discussions of potential violations of the Sure-Thing Principle). Thus, in order to argue that practical rationality restricts rational agents' attitudes in a substantive way, one must be committed to the claim that there is some limit to the individuation of the alternatives with respect to which practical rationality is defined.

A general challenge normative decision theory faces, then, is to find a way to draw the line between admissible individuations of the alternatives and inadmissible ones. Facing this general challenge is particularly important for advocates of LP who wish to argue that the use of nudges does not limit people's freedom to choose any of the alternatives that would have been available to them if nudges had not been used.

In the absence of a justified method for drawing this line, even paradigmatic cases of nudging cannot be taken to fall under T&S's own definition. To borrow an example from Selinger and Powys (2010), even describing the decision at Schiphol airport to paint a little image of a black fly on all the urinals as a nudge might be unjustified since it seems plausible that rationality does not forbid one from having a strong preference to avoid urinating on images of living creatures. Thus, by painting the fly images, the airport authorities did eliminate an alternative from the choice set (i.e. using an image-free urinal).

Admittedly, and as Sunstein (see his 2015b) argues (in a different context), this *conceptual* difficulty does not seem to threaten governments' ability to use nudges, since there is often wide agreement in the case of many specific interventions that they indeed only interfere with the choice architecture and do not change the alternatives themselves (even though no general criterion for distinguishing between the two is identified). However, my aim here is not to defend (or argue against) the governmental use of nudges. Rather, it is to examine how it is possible to take nudges as the expression, at the level of policy, of a picture of the social world that has ideological content. In order to do so, it is crucial to meet the conceptual challenge of finding a criterion or a method for drawing the line between the choice architecture and the alternatives in the choice set. This is so since, as we just saw, the core idea of LP, namely the possibility of influencing people's behaviour without limiting their freedom of choice, depends on the in-principle possibility of drawing such a line.

b. Meeting the conceptual challenge

Hurley began her investigation in “Natural Justice” with the general problem, discussed above, of the individuation of alternatives. This problem can be approached from two perspectives: the third-person and the first. The former focuses on attempts to understand whether the behaviour of another person is rational, and the latter focuses on attempts to understand whether one’s own behaviour is rational.

These two points of view on the problem of the individuation of alternatives can plausibly be taken to constitute two different *problems*. The first, which is associated with the third-person perspective, can be understood as a *methodological* problem for social scientists (see for example Tversky 1967), while the second, that associated with the first-person perspective, can be understood as a purely *normative* problem (see for example Broom 2009, chapter 5). However, Hurley disagreed. She insisted that there is in fact only one problem, which may be approached from two different points of view, namely a problem of *interpretation*. The reasons why Hurley believed this are important for us. In order to expose them, it is best to concentrate on a concrete example (the following exposition differs from Hurley's more systematic and detailed – albeit much longer – one, but I believe it expresses Hurley's main insight).

Let us start with the third-person perspective. An observer observes a series of choices made by an agent and has to decide whether this pattern of choices is rational. In order to do so, he must first individuate the set of alternatives amongst which the agent can choose. As we saw, different individuations may lead to different judgements regarding the rationality of the agent's behaviour. Suppose, for example, that the observer, Ogg, observes that the agent, Amal, chooses coffee over beer when the two are presented to her at 9.00 a.m., chooses beer over wine when the two are presented to her a few minutes later, and chooses wine over coffee when the two are presented to her at 11:35 a.m.

Ogg has to decide whether or not this pattern of choices made by Amal constitutes an instance of intransitivity. In order to do so, he has to decide whether "beer at 9.00 a.m." is the same alternative as "beer at 11:35 a.m.". A natural criterion to adopt here is Amal's own opinion. Rational choice theorists often argue that instrumental rationality deals only with structure and not with content. It says nothing, according to conventional wisdom, about what to value or what to prefer; it only forbids certain *patterns* of preferences. Thus, it seems to be against the Humean spirit of rational choice theory to impose on Amal an external criterion for the individuation of the alternatives in the choice she faces. If it matters *to Amal* whether it is 9.00 a.m. or 11:35 a.m. when she is drinking her beer, then Ogg must take these to be two different alternatives when he evaluates Amal's behaviour.

This answer, however, leads us to the first-person perspective. When Ogg turns to Amal and asks her whether she distinguishes between "beer at 9.00 a.m." and "beer at 11:35 a.m.", Amal stands in front of the following dilemma: she can either take the two descriptions to constitute two different alternatives, and take her previous choices to be in line with the transitivity of choice demand, or she can deny that the two descriptions constitute two different alternatives and so take her previous choices to be irrational (thus, she might be motivated to change her manifested preferences). Sometimes Amal might find this dilemma easy to solve, but – I take this to be revealed by reflection – not always.

Now, let us concentrate on cases in which Amal does not find this dilemma easy to solve. What is the right way to describe Amal's attitude to the dilemma in these cases?

Is it appropriate to describe Amal as being *uncertain* about whether "beer at 9.00 a.m." is a different alternative from "beer at 11:35 a.m.", or is it more appropriate to describe her as being "torn" between the two horns of a dilemma? Is solving the dilemma, in other words, a matter of forming a correct belief or a matter of making a decision?

If it is a matter of a decision, then the conclusion must be that Amal can *make* any pattern of choices rational *by will*. If it is a matter of belief, then we are back to the third-person point of view, only now Amal occupies both the role of the observer and that of the agent.

There seems to be something intuitively right and something intuitively wrong about each one of these two directions. First, if Amal – having gone through an appropriate process of deliberation – declares that she does value drinking a beer at 9.00 a.m. differently than she does drinking a beer at 11:35 a.m., it seems extremely unintuitive to argue that Amal can be *wrong* about that.[7] Thus, taking her declaration to express a decision that – once taken – *makes it the case* that the two descriptions constitute two different alternatives, seems appropriate. However, if the declaration expresses *nothing but a decision*, it is unclear which mental resources are used by Amal when she reflects on the problem. If Amal can just make her claim true by will, why is the reflection needed in the first place?

Intuitively, when Amal reflects on the dilemma, what she does is look for information and assess information she already has with the aim of arriving at a true judgement. Thus, it seems that Amal's declaration should be understood as expressing both a belief she holds and a decision she has made. Similarly, the mental process Amal goes through

[7] Whatever is the criteria for the "appropriateness" of the process of deliberation Amal goes through (as long as they do not include the trivial criterion according to which a process of deliberation is appropriate iff it leads to the correct judgement).

when she reflects on the matter should be seen as a process that has features of both practical reasoning (which aims at a decision) and theoretical reasoning (which aims at a true belief). The key for solving the problem of the individuation of the alternatives seems to lie, then, in correctly characterizing this mental process.

Hurley suggested such a characterization. According to Hurley, the mental process Amal goes through is a process of "self-interpretation and self-determination". Amal's reasoning aims at finding out what the right individuation is and, at the same time, plays a constitutive role in the determination of this individuation.

Here is the main idea. Eligible distinctions, ones that are allowed by the correct individuation of the alternatives, are ones that are supported by "human nature" (see Hurley 1989, chapter 5.3), when this notion is understood as being sensitive to both physical (including biological) and social constraints (and thus can change when social structures changes – I return to this point below). When an agent reflects on issues of the eligibility of distinctions between alternatives, she is engaged in a process of forming beliefs regarding human nature. A complete theory of which distinctions are eligible and which are not is part of what Hurley calls an "ethic".

An ethic, in Hurley's sense of the term, is a theory of *practical rationality* that specifies the substantial substantive (as opposed to purely structural) demands of rationality. Most importantly for our discussion here – an ethic fixes the set of eligible distinctions, the set of distinctions supported by human nature. Inspired mainly by Davidson, Hurley adopts a coherence account of ethics. According to Hurley, the right theory of substantive rationality is the theory of substantive rationality that "best displays as coherent the relationships among specific reasons for action" (Hurley 1989, p. 225). Hurley not only postulated the existence of such a theory (in the way some coherence accounts do), but also forcefully argued for it (see Hurley 1989, chapters 13-14). As my aim here is not to defend Hurley's view from various criticisms, but rather to demonstrate how her suggestion can help proponents of LP meet the challenges they face, I will not present her argument here.

It is nevertheless important to understand the significance of the appeal to coherence in Hurley's account. The correct individuation of the alternatives is determined by human nature. Part of human nature, however, is to care about self-determination (Hurley goes even further, suggesting that it might be "one of the distinguishing features of persons, or one of the distinctive functions of human beings as rational animals" [Hurley 1989, p. 114]). However, to say that self-determination is part of human nature is to say that the best theory of substantive rationality, the one that best displays as coherent the relationships between different reasons for action, says that self-determination is an important value.

By "self-determination" Hurley refers to the mental process by which a person fixes his first-order attitudes (and specifically his preferences). As explained, this process is that of finding an ethic. Thus, the process of finding an ethic is both a process of self-interpretation, a process of looking for the most coherent theory of the relationship between one's own reasons for action, and of self-determination, that is, of fixing – through one's deliberation – one's values.

What we have arrived at is a worked-out suggestion for how to meet the first challenge presented in the previous section, the challenge of finding a criterion for the individuation of alternatives. According to Hurley, finding such a criterion is a substantive matter: it is part of one's ethic and any plausible ethic requires that *one be engaged in an active process of deliberation that aims at finding out one's ethic*. This process is a process of looking for the most coherent interpretation of oneself. The most coherent interpretation is determined by both one's first-order attitudes and one's higher-order attitudes. It is determined by the things one cares about and by the things one wants to care about, by the kind of person one is and by the kind of person one wants to become.

In order to see things more clearly it might help to return to the example given in the previous section. Suppose a public servant has to decide whether or not, by fixing the default choice for about-to-get-married couples as to "signing a prenuptial agreement", the government would eliminate an alternative from the choice set available to a specific future bride, Bella (or only change the choice architecture of the choice Bella and her future husband face). According to Hurley's suggestion, the answer to this question lies in Bella's ethic, and thus to answer the question the public servant has to look for Bella's ethic, look, in other words, for the most coherent interpretation of Bella as an agent who weighs different reasons for actions against each other.

Now, on the one hand, this interpretation is not necessarily the one Bella herself adopts. Bella might be wrong about what her ethic is. She might, for example, go through a biased process of deliberation on this question. On the other hand, however, since this interpretation must be determined through a process in which Bella *herself* is looking for her ethic, the answer to the question crucially depends on Bella's deliberation. Specifically, if Bella does not take part in the process of deliberation, the interpretation chosen by the public servant is bound to be the wrong one to adopt (since it is arrived at by a process that was not sensitive to the value of self-determination).

It is, of course, true (as an anonymous referee commented) that this theoretical response to the conceptual challenge cannot always be translated into a practical suggestion for how to determine whether or not a given intervention constitutes a nudge with respect to a specific individual.[8] However, in this paper my focus of attention is not the implementation of specific nudges, but rather the search for general guidelines regarding the institutional design that supports a justified implementation of nudges.

Any such institutional design will, surely, fail with respect to specific interventions and specific individuals in many cases (due to both the epistemic limitations of public agencies and the psychological diversity between individuals). This is true not only when it comes to the identification of interventions as nudges (which is what the conceptual challenge is about), but also when it comes to choosing the types of behaviours these interventions aim to promote. Nudges are supposed to promote people's welfare, *as judges by themselves,* but different people have different judgements regarding what welfare consists of. Thus, any institutional arrangement in the spirit of LP will, in some cases, fail to implement justified nudges with respect to some citizens. It does not follow, however, that all such institutional arrangements are equally good (or bad) in implementing nudges. Some institutional arrangements are better than others and the road for designing the best one must go through a better understanding of the conceptual foundations of LP.

The Hurlyian response above does point in a general direction in the search for guidelines for justified procedures for making high-order decisions regarding the implementation of nudges. This will be discussed in section 4. However, before we can move to the political level, we must, first, review the ethical level.

[8] Although in some cases it can. This seems to be the case, for example, when it comes to "clinical nudges" in healthcare (i.e. nudges that health practitioners use to influence the decisions of specific patients). See for example the discussion in Avitzour et al. (2019).

3. The Ethical Challenge

a. Presenting the ethical challenge

Several different ethical criticisms have been directed at LP (see Schmidt and Engelen 2020 for a review). It would be fair to say, however (and this is also reflected in the public view of nudging; see Sunstein 2016 for evidence), that the most troubling is the claim that nudges intrude on personal autonomy (see Blumenthal-Barby and Naik 2015 and Grune-Yanoff 2012 for useful discussions).

Of course, what exactly personal autonomy is is a controversial matter. As Hacker (2018) nicely explains, different notions of personal autonomy give rise to different versions of the criticism from autonomy (and these criticisms might be specifically challenging for different types of nudges). However, I take it that all these versions of the criticism aim to capture the same strong intuition that often arises when different nudges are considered. In the next sub-section, Hurley's own notion of personal autonomy will be presented and it will be argued that adopting this notion gives rise to a version of LP that not only does not conflict with personal autonomy but also treats autonomy as the main value that grounds it.

It is possible to point to three main lines of response adopted by advocates of LP to the criticism(s) from personal autonomy, even without committing ourselves, at this stage, to one specific explication of the notion. First, it is often argued that not all types of nudge conflict even intuitively with personal autonomy, and some types of nudge clearly support personal autonomy (specifically, nudges such as the "cooling down period" nudge, which trigger conscious deliberation, seem to belong to this group; see Barton and Grune-Yanoff 2015 for discussion).[9]

There is no doubt that this response is effective in the type of cases in which it is applicable, but this still leaves most of the nudges discussed in the literature vulnerable to the criticism.

Second, it is argued that even in cases in which the interference with autonomy is real, it should be weighed against the contribution of the specific intervention in question to people's welfare and should – sometimes – be tolerated (see for example Sunstein 2015b). Again, there is no doubt that this response is effective, but it is effective even in cases of outright classical paternalism. Thus, for the purpose of constructing conceptual and ethical foundations for LP, this response is uninteresting.

Third, and most importantly, it has been argued that whenever nudges do interfere with personal autonomy, personal autonomy is compromised anyway. This response is based, I take it, on the following (often implicit) inference: if designing a certain feature of the choice architecture in a specific way amounts (or leads) to intruding on personal autonomy, then designing the same feature of the choice architecture in a different way amounts to intruding on personal autonomy as well. Thus, to the extent that nudges do intrude on autonomy, this intrusion is unavoidable (see, for example, Thaler and Sunstein 2008 and Sunstein 2015b for this response).

[9] Schmidt (2020) adopts a more radical line of defense according to which all (or at least most) nudges do not interfere with autonomy. However, in order to support this claim, he adopts a very pluralistic conception of rationality, ecological rationality. I think Schmidt's argument work for those who adopt his conception of rationality. Furthermore, in her latter work (Hurley 2003 and 2005) Hurley herself adopted a similar notion of rationality. However, I also think that Hurley's analysis in *Natural Reasons* has the resources to offer a more general response to the challenge from autonomy, a response that holds also for stricter notions of rationality.

It is not at all clear to me that this inference is valid (whether it is valid depends on the specific explication of the notion of personal autonomy adopted). However, to the extent that it is valid, while the response can serve as a good way to justify the use of specific nudges in specific cases, I find it inadequate for the purpose of supplying LP with solid foundations.

If it is indeed true that, under the contemporary social and institutional order, intrusions on personal autonomy are unavoidable in a wide range of cases, the correct response on the level of political ideology should be much more radical. It should be either to question the basic centrality of personal autonomy in one's notion of liberal democracy (in the way Conly 2012 does, for example) or to keep autonomy's central ideological place but support a radical change in the institutional arrangement of democratic societies (such as the one that Marcuse 1964 calls for, for example. Notice that Marcuse's main criticism of the institutional structure of contemporary democratic societies was based on his claim that this structure intrudes on personal autonomy. Notice further that although Marcuse did not have at his disposal the rich psychological literature that T&S draw upon to support their claims, his claim is based on observations very similar to those T&S make).

Admitting that autonomy is often compromised and treating this as an unavoidable state seems a peculiarly pessimistic basis for a political ideology adopted by someone who takes personal autonomy to be a central moral value. Thus, for non-radicals who are unwilling to abandon personal autonomy as a central moral value, one ethical challenge LP faces seems to be that of presenting an explication of personal autonomy according to which the use of nudges[10] is not only compatible with personal autonomy but is also the tool through which autonomy becomes possible.

b. *Meeting the ethical challenge*

To practice one's autonomy, according to Hurley, is to be engaged in a process of self-determination in the way explained in the previous section. According to Hurley, this view of personal autonomy stands in sharp contrast to (what we call today) hierarchical views of personal autonomy, views that understand personal autonomy as the fixation of one's first order preferences and desires by one's (fixed) higher-order preferences and desires (e.g. Frankfurt 1988, Dworkin 1988). Hurley describes such approaches to personal autonomy as fundamentally subjective: the criterion for whether an agent is an autonomous agent is the agent's own high-order preference.[11] In contrast, Hurley describes her own view as objective: whether an agent is autonomous does not depend on the preferences she ends up having or on how they match some fixed privileged set of her preferences (e.g. her extended preferences). Rather, it depends on the process through which she has arrived at those preferences.

The objective-subjective distinction is suggested by Hurley herself. However, and as an anonymous referee commented, this is not the terminology used in the contemporary literature on personal autonomy. Although Hurley’s notion of personal autonomy cuts across some of the distinctions used today in this literature, I think it is fair to describe her account as coherentist (because the process of self-interpretation and self-

[10] The use of *all* nudges, not only those which are specifically designed to trigger deliberative faculties, such as "the cooling down" nudge, but also those nudges which are system 2 based.

[11] Thus, argues Hurley, they are vulnerable to similar objections to the ones directed at the claim that just any action in light of one's preferences is autonomous. I will not present Hurley's argument against such approaches here – it can be found in Hurley 1989, chapter 6.

determination aims to arrive at the most coherent theory of oneself as an agent that correctly responds to her reasons for action), historical (because it is not just the end-state of the process that matters, but rather an active engagement in the process by the agent herself is required), but, at the same time, externalist (because the end state is the most coherent theory of the different reasons for action the agent has and these reasons are external to the agent).

Hurley's notion is also fundamentally relational (although she does not use this term) but, as I discuss below, the emphasis she gives to the social nature of autonomy stems from different sources than those that stand at the basis of other relational accounts of personal autonomy (see for example the papers is Mackenzie and Stoljar 2000, Westlund 2009 and the discussion in Christman 2004).

The process, the engagement in which constitutes, according to Hurley, being autonomous is the process of self-interpretation and self-determination. This process, we saw, must involve direct engagement with questions regarding the eligibility of distinctions, i.e. regarding which patterns of preferences are rational.

We can see now that, according to Hurley, the phenomenon of irrational behaviour is *not only not a threat to autonomy, but rather the key to exercising it*: one becomes autonomous only through one's engagement with questions about one's own rationality.[12] Bella can make an autonomous decision regarding whether to sign a prenuptial agreement only by engaging in a process of deliberation that aims at finding whether being influenced by whether or not the default option is "sign a prenuptial

[12] This is so, it is important to emphasis, irrespectively of one's notion of rationality. As explained in the previous section, as long as one's notion of rationality is not the trivial one that put no restrictions on rational behavior, questions about the eligibility of distinctions will arise.

agreement" is rational. If people's behaviour was always rational, autonomy could not be exercised in this way.

This point of view nicely explains the relations between rationality and autonomy. Rationality, as we saw, can restrict behaviour only given a commitment to certain substantive restrictions on the eligibility of distinctions. Such a commitment can be arrived at only through a process of self-determination and self-interpretation, and this process is the way in which people exercise their autonomy. Thus, being autonomous is a condition for being rational (because rationality requires a given individuation of the alternatives, which can be fixed only through exercising one's autonomy), and one can be autonomous only by being engaged in a process that aims at arriving at a rational choice. Without the possibility of failing to reach this aim, that is, without the possibility of acting irrationally, this process cannot take place.

Recall that the challenge from autonomy was not only to present an explication of the concept that allows us to treat people as autonomous while recognizing the widespread phenomenon of irrationality, but also to understand the use of nudges as a way to make autonomy possible pace this phenomenon. To see how Hurley's approach can meet this challenge we must discuss the *social nature* of the process of self-determination and self-interpretation through deliberation that was just presented.

Two different types of consideration – constitutive and epistemic – make it the case that this process (which, as explained, is the process through which agents exercise their autonomy) must have a social nature. From an epistemic point of view, notice first that some of the reasons for action for a given individual are (or are grounded in) reasons for action of other people. For example, some (but not all) of Bella's reasons for signing or avoiding to sign a prenuptial agreement are facts regarding Bella's partner's reasons for signing or avoiding to sign such an agreement. Since Bella's ethic, the theory of practical rationality Bella ought to adopt, is the theory that best displays Bella's reasons for action as coherent, since some of these reasons are grounded in Bella's partner's reasons for action, and since Bella's partner is (usually) in a unique position to know what these latter reasons are, therefore when Bella is engaged in a process of deliberation that aims at finding her ethic she should consult her partner. The same is true, of course, regarding any other person whose reasons for action matters (to some extent) to Bella. Moreover, regarding some aspects of her life, Bella's partner (or other people) may be in a better position than Bella herself to assess whether or not – and to what extent – certain features of her life constitute reasons for certain actions she might take.

Second, when making judgements regarding the level of coherence of different possible theories, Bella should consult with people who are in a better epistemic position than she to assess these levels of coherence. For example, Bella might ask herself which one of the following two interpretations of her initial attitudes regarding signing a prenuptial agreement is more coherent. According to the first interpretation, Bella cares about whether or not signing a prenuptial agreement is set as the default option, and has a consistent set of preferences that ranks signing such an agreement when it is set as the default option above not signing it (whether or not it is set as the default option), which is ranked above signing it when it is not set as the default option. According to the second interpretation, Bella does not care about whether or not signing the agreement is set as a default option and has (at the starting point of her deliberation) inconsistent attitudes: she prefers signing an agreement to not signing it when it is set as the default option and not signing it to signing it when it is not set as the default option, even though she takes as ineligible a distinction that is based on whether or not signing the agreement is the default option.

Now, the answer to the question as to which of these two interpretations is more coherent is partly determined by considerations that may be unavailable to Bella herself, but are available to other people. For example, scientific evidence that relates Bella's tendency to stay with the default option (whatever it is) in the case of the possibility of signing a prenuptial agreement to psychological mechanisms that are responsible for other behavioral tendencies that Bella herself takes to be irrational (such as a tendency to stay with the default option – whatever it is – in cases in which *it is clear* to Bella that she does not care what the default option is), supports the second interpretation, the one that takes Bella's attitudes before she starts her deliberation to be irrational. On the other hand, scientific (or other) evidence that relates Bella's tendency to stay with the default option in the case of a prenuptial agreement to a general preference to avoid legalizing her romantic life supports the first interpretation.

As mentioned, epistemic considerations are not the only ones that give rise to the social nature of the process of self-interpretation and self-determination. Constitutive considerations are also relevant here. The question of which distinctions are eligible is understood by Hurley, as we saw, as a question about human nature. An eligible distinction is one that is natural for humans to make. Now, on the one hand Hurley insisted that we must allow for cases in which some people make distinctions that are unnatural (as, if such cases did not exist, no non-trivial condition of rationality could be formalized). On the other hand, however, what human nature consists of is determined, as we saw, by the process of self-interpretation and self-determination itself. Thus, in order for this process to be both constitutive of human nature and one that can make some distinctions made by some people eligible, it must be social: it is determined by a social process in which each individual takes an active part, but which is also external to each individual.

One important implication of the (necessarily) social nature of the process of self-interpretation and self-determination is that one needs others in order to exercise one's autonomy. It is impossible, we have just seen, to go through an adequate process of self-interpretation and self-determination without relying on other people. Thus, rather than being a threat to personal autonomy, giving society an *adequate* role in determining individual preferences is a necessary condition for individual preferences to be autonomous. Specifying guidelines for an institutional design that will promote such a role for society is the political challenge which I discuss next.

The political challenge

a. Presenting the political challenge

In the last chapter of their book, T&S write:

> The twentieth century was pervaded by a great deal of artificial talk about the possibility of a 'third way'. We are hopeful that libertarian paternalism offers a real Third Way – one that can break through some of the least tractable debates in contemporary democracies. (Thaler and Sunstein 2008, p. 253)

At the same time, however, T&S (as well as other advocates of LP) insist that the policy tools recommended by LP can be (and are) adopted by both liberals and conservatives. This seems odd. If all that LP does is suggest interventions that no one has a reason to oppose, in what sense does it constitute an alternative to traditional political ideologies?

Now, one might (and in informal communication many philosophers do) argue that LP should indeed not be treated as a comprehensive political ideology on par with social democracy, libertarianism or conservatism. Rather, it should be understood in a much more modest way: it is merely a commitment to the approval of the use of nudges as uncontroversial policy tools. While such a position is clearly at odds with the aspirations of T&S as expressed in the quote above, it is certainly consistent to support the use of nudges (or some nudges) by governments without getting engaged in the project of constructing LP as a comprehensive political ideology.

However, even those who adopt such a modest commitment to LP still have to answer high-order questions about the processes by which, and the institutional design in which, decisions about implementation of nudges are taken.

Moreover, at least some advocates of LP – in academic publications, in popular media and in public administration – seem to take LP to be more than the mere commitment to the use of nudges as uncontroversial policy tools (even if they believe it falls short of a comprehensive political ideology). I think they have good reasons for this.

Independently of the specific policy recommendations associated with it, LP is characteristically rooted in the research in the behavioural sciences (especially cognitive psychology and behavioural economics) that explores the various ways in which people's decision making and reasoning capacities deviate from the ideal of perfect rationality. Taking the image of Homo sapiens that emerges from this body of research seriously means more than simply designing policy tools that either take advantage of or help people overcome their cognitive limitations. As discussed in the previous sub-sections with respect to the ideas of personal autonomy and rational agency,[13] the challenges, as well as the opportunities, which these empirical findings pose to political theorists are much wider and deeper.

Adopting the modest interpretation of LP amounts to ignoring these challenges and not taking advantage of these opportunities. For those who take these empirical findings seriously, these challenges ought to be faced and these opportunities ought to be taken advantage of. At the level of political theory, the task such advocates of LP face is to address these challenges in a way that naturally leads to the use of nudges, not as local solutions to local problems, but rather as an expression of a unified picture of the social world that has both descriptive and normative elements (even if it is one that falls short of a comprehensive political ideology). This, I will now argue, can be achieved using Hurley's political theory.

b. Meeting the political challenge

What is society's adequate role in determining an individual's preferences? Well, as we have seen, society should, on the one hand, help the individual overcome the biases he is subject to in case such biases exist; but on the other hand, society cannot alienate the individual from the process of deliberation altogether. The social nature of the process of self-interpretation and self-determination relies on what Hurley calls "a cognitive division of labor", on making use of experts' judgements without fully transferring judgment to another person. This last is so, as was explained, because it is in human

[13] However, the same is true with respect to other concepts that play a central role in most political ideologies, such as welfare, responsibility and so forth. My discussion in this essay touches mainly on autonomy and rationality. However, Hurley's political theory is much wider and cover these other concepts as well.

nature to seek self-determination and complete deference to other people contradicts self-determination.

Now, implementing a nudge is a way to help people overcome the biases they are subject to by using experts' judgements (remember Bella's reliance on scientific evidence regarding the psychological mechanism responsible for her choices). However, in order for the implementation of a specific nudge to foster rather than threaten the personal autonomy of subjects, these subjects must not be completely alienated from the process through which the nudge is implemented.

This leads to conclusions on the political level. Using nudges (even those nudges that clearly take advantage of people's biased cognitive capacities) as a policy tool can – if done correctly – not only be compatible with respecting people's personal autonomy but also might be the only way to enable them to fully exercise it. In cases in which people are bound to act irrationally, in cases in which they either fail to go through a process of deliberation, or go through a biased process, they cannot fully exercise their autonomy without external help. Nudges can enable them to do this, *provided that the way these nudges are implemented and the way the decision to use them is made is subject to public debate.* In such a debate, social scientists should have a privileged position regarding matters that fall under their areas of expertise, but they cannot be the only participants. If they were, the process of self-determination and self-interpretation they seek to support using their expertise would not be possible because the "self-determination" part of it would be compromised.

These two conditions on the appropriate way to implement nudges that we have inferred from Hurley's position – the one that gives a privileged position to scientists' positions regarding matters that fall within their areas of expertise, and the one that demands that people are not alienated from public debate regarding which nudges to implement and how – reflect two general principles regarding the appropriate institutional design for liberal democracies that Hurley herself pointed to. Referring to a community that is involved in the kind of process of self-determination and self-interpretation that she described in her book, Hurley writes:

> …even in the absence of knowledge of the truth, beliefs about certain kinds of issues, which have been formed in certain ways or under certain circumstances, may be known to be debunked, as the result of self-deception, wishful thinking, prejudice, deceit, propaganda, advertising, or some other kind of deliberate manipulation, a kind of sour grapes reaction, illusion, common inferential error, etc. The members of such a community should seek to exploit this possibility by designing procedures and institutions that effect a division of epistemic labor in accordance with generalizations about the circumstances in which beliefs about certain types of issues are debunked, or likely to be. *First,* they should seek to divide authority on various kinds of issues among institutions and procedures not so much in accordance with positive expertise about what should be done, which may be very hard to identify in the absence of knowledge of the truth about it, but so as to avoid relying on debunked (or debunkable…) beliefs. *Second*, they should actively foster the capacity for deliberation and the formation of undebunked beliefs. (Hurley 1989, p. 326)

In this quotation, Hurley refers to beliefs, not choices. However, as should by now be clear, it follows from Hurley's notion of rationality that a rational agent's choices are an expression, not of a set of fixed preferences she somehow comes equipped with, but rather of a set of all-considered judgements about what is to be done that were formed in an appropriate process of self-interpretation and self-determination. Thus, the first principle Hurley points to in the quotation above corresponds to the recommendation to give social scientists a privileged role regarding matters that fall under their field of expertise in public debates regarding which nudges to implement and how, and the second principle to which she points corresponds to the recommendation to make every judgement a matter of open public debate.

Thus, the response to the second challenge for LP outlined in the previous section, the challenge from autonomy, which Hurley's philosophy points to, is very different from Thaler and Sunstein's original response. The point is not that LP allows for violations of autonomy only in cases where autonomy is compromised anyway, i.e. in cases in which people are subject to systematic biases. On the contrary, the response is that those cases – cases in which people are subject to systematic biases – are exactly those in which personal autonomy can be exercised, and it can be exercised only through a process in which society plays a crucial role.

Hurley takes autonomy to be a value that comes in degrees. One can be more or less autonomous. Sometimes, Hurley believes, it is appropriate to give up on some autonomy in order to promote other values (such as welfare or negative liberty). However, although it does not have a moral priority over other values, the value of autonomy has a distinctive role in Hurley's conception of liberal democracy in virtue of its structural relations to other values. One exercises one's autonomy partly by making decisions regarding how much autonomy one is willing to sacrifice for other values. A political system that aims at giving its members the best conditions to exercise their autonomy is thus also – out of a conceptual necessity – the one that leads to the best compromises between all the relevant competing values.

Now LP, as noted, is characteristically rooted in the findings in social science that point to those systematic biases in people's decision-making capacities *that arise in certain types of cases*. This is an alternative image of rational agency to those adopted (often implicitly) by traditional views that are either committed to a very weak conception of rational agency (according to which rational requirements are minimal and, thus, most of the time people are rational) or a very strong one (according to which rational requirements are very demanding and, thus, most of the time people act irrationally). Hurley’s notion of rational agency is an attempt to save the image of Homo sapiens as a rational animal pace the findings in social science in which LP is rooted. It is a conception of rational agency that cannot be described as either weak or strong since it is not committed to any external set of criteria relative to which the rationality of people is evaluated. Decisions are rational, or not, only relative to a given individuation of the alternatives and the correct individuation is determined through a process of self-interpretation and self-determination.

This conception of rational agency gives rise to Hurley's conception of personal autonomy; a conception, as was demonstrated, which succeeds in meeting the most pressing ethical challenge for LP, namely the challenge from autonomy. Thus, it seems natural for LP to adopt Hurley's conception of liberal democracy. By so doing, proponents of LP can meet all three challenges presented in the previous section – the conceptual challenge, the ethical challenge and the political challenge – in a unified way.

References

Avitzour,D, Barnea, R, Avitzour, E, Cohen, H & Nissan-Rozen, I. (2019), Nudging in the clinic: he ethical implications of differences in doctors' and patients' point of view, *Journal of Medical Ethics,* 45, 183-89.

Barton, A. and T. Grune-Yanoff. (2015), From libertarian paternalism to nudging – and beyond, *Review of Philosophy and psychology*, 6:3, 341-59.

Bhattacharyya, A. Pattanaik, K. P. and Xu, Y. (2011), Choice, Internal Consistency and Rationality, *Economics and Philosophy,* 27, 123-49.

Blumenthal-Barby, J.S. and Naik, A. D. (2015), In Defense of Nudge-Autonomy Compatibility, *American Journal of Bioethics,* 15:10, 45-7.

Bovens, L. (2009). The Ethics of Nudge. Pp. 207–20 in Preference Change, edited by T. Yanoff-Grune and S. O. Hansson. Dordrecht: Springer.

Broome, J. (1991), *Weighing Goods*, Cambridge, Mass.: Basil Blackwell.

Broome, J. (1999), *Ethics out of Economics,* Cambridge, UK: Cambridge University Press.

Christman, J. (2004), Relational Autonomy, Liberal Individualism, and The Social Constitution of Selves, *Philosophical Studies*, 117: 1/2, 143-64.

Conly, S. (2012), *Against Autonomy: Justifying Coercive Paternalism,* Cambridge University Press.

Dworkin, G. (1988), *The Theory Practice of Autonomy,* Cambridge University Press.

Ekstrom, W. L. (1993), A Coherence Theory of Autonomy, *Philosophy and Phenomenological Research,* 53 (3), 599-616.

Engelen, B. (2019), Nudging and Rationality: What is there to worry?, *Rationality and Society,* 31 (2), 204-32.

Frankfurt, H. (1988), *Freedom of the will and the concept of a person,* in *The importance of what we care about,* Cambridge University Press.

Fumagalli, R. (2020), On the Individuation of Choice Options, *Philosophy of the Social Science,* 50:4, 338-65.

Grune-Yanoff, T. (2012). Old wine in new casks: libertarian paternalism still violates liberal principles. Social Choice and Welfare 38: 635-645.

Hacker, P. (2018), Nudging and Autonomy: A Philosophical and Legal Appraisal, in *Handbook of Research Methods in Consumer Law, Ed. Edward Elgar.*

Hausman, D. M. and Welch, B. (2010), Debate: To Nudge or Not to Nudge, *The Journal of Political Philosophy,* 18:1, 123-36.

Hurley, S. (1989), *Natural Reasons: Personality and Polity,* Oxford University Press.

Hurley, S. (2003), The limits of individualism are not the limits of rationality, *Behavioral and Brain Sciences,* 26:2, 164-5.

Hurely, S. (2005), Social Heuristics that make us smarter, *Philosophical Psychology,* 18:5, 585-612.

Lepenies, R., and Malecka, M. (2015), The institutional consequences of nudging — nudges, politics, and the law, *Review of Philosophy and Psychology,* 6:3, 427-37.

Mackenzie, C. & Stoljar, N. ed. (2000), *Relational Autonomy,* Oxford University Press.

Marcuse, H. (1964), *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society,* Beacon Press.

Mills, C. (2013), Why Nudges matter: A Reply to Goodwin, *Politics,* 33:1, 28-36.

Mills, C. (2015), The Heteronomy of Choice Architecture, *Review of Philosophy and Psychology,* 6:3, 495-509.

Saghai, Y. (2013a), Salvaging the Concept of Nudge, *Journal of Medical Ethics,* 39:8, 487-93.

Saghai, Y. (2013b), The Concept of Nudging and its Moral Significance: A Reply to Ashcroft, Bovens, Dworkin, Welch and Wertheimer, *Journal of Medical Ethics,* 39:8, 2012 -2021.

Schmidt. T. A. (2017), The Power of Nudge, *American Political Science Review*, 111 (2), 404-17.

Schmidt, T. A. (2020), Getting Real on Rationality – Behavioral Science, Nudging, and Public Policy, *Ethics,* 129, 511-43.

Schmidt, T. A. & Engelen, B. (2020), The Ethics of Nudging: An Overview, *Philosophy Compass,* **https://doi.org/10.1111/phc3.12658**

Selinger, E. and Powys Whyte, K. (2010), Competence and Trust in Choice Architecture, *Knowledge, Technology and Policy,* 23:3, 461-82.

Sunstein, C. R., & Thaler, R. H. (2003), Libertarian paternalism is not an oxymoron, *The University of Chicago Law Review*, 70:4, 1159–1202.

Sunstein, C.R. and Thaler, R. (2008), Nudge: Improving Decisions about Health, Wealth, and Happiness. New Haven CT: Yale University Press.

Sunstein, C.R. (2015a), Nudges, agency, and abstraction: A reply to critics, *Review of Philosophy and Psychology*, 6:3, 511-29.

Sunstein, C.R. (2015b), Nudging and choice architecture: ethical considerations, *Discussion Paper No. 809*, Harvard Law School.

Sunstein, Cass R. (2016), *The Ethics of Influence*, Cambridge University Press.

Tversky, A. (1967), Additivity, Utility, and Subjective Probability, *Journal of Mathematical Psychology,* 4:2, 175-201.

Westlund, C. A. (2009), Rethinking Relational Autonomy, *Hypatia,* 24:4, 26-49.

Wilkinson, T.M. (2013), Nudging and manipulation, *Political Studies*, 61:2, 341-355.