Philosophy of Management (2020) 19:45–68
https://doi.org/10.1007/s40926-019-00117-w

# Nudging and Participation: a Contractualist Approach to Behavioural Policy

Johann Jakob Häußermann[1,2]

Published online: 4 July 2019


**Abstract**
As behavioural economics reveals, human decision-making deviates from neoclassical assumptions about human behaviour and people (often) fail to make the 'right' welfare-enhancing choice. The purpose of Sunstein and Thaler's concept of 'nudge' is to improve individual welfare. To provide normative justification, they argue that the only relevant normative criterion is whether the individual is 'better off *as judged by themselves*', so that the direction in which people are to be nudged is defined by their own preferences. In light of behavioural findings, however, people's choices do not provide a sound basis for eliciting preferences and thus for assessing welfare. In this paper, I aim to challenge Sunstein and Thaler's normative view, arguing that it is unreasonable to rely on conventional welfare economics, particularly considering the given context. Sunstein and Thaler depend on an approach of 'preference purification' which assumes informed, latent, and true preferences: As a result they face crucial methodological, epistemological, and practical objections, and cannot show how their approach enhances individual welfare. By building on the concepts of R. Sugden and C. Schubert, I develop an alternative normative framework for behavioural public policy, based on a contractualist perspective in which people may consent to collective choice rules in order to align future behaviour with values, to achieve particular goals or to preserve personal integrity. Individual consent and citizens' participation and deliberation are crucial to this approach. This contractualist approach may provide a normative justification for behavioural public policy, and help to reconcile behavioural and normative economics.

**Keywords** Behavioural policy · Nudge · Welfare economics · Libertarian paternalism · Participation

✉ Johann Jakob Häußermann
johann-jakob.haeussermann@iao.fraunhofer.de

[1] Fraunhofer Institute for Industrial Engineering IAO, Center for Responsible Research and Innovation, Stuttgart, Baden-Württemberg, Germany

[2] TUM School of Governance, Technical University Munich, Munich, Germany

## Introduction

Based on the emergence of behavioural economics, C. Sunstein and R. Thaler developed the concept of 'nudge'; policy designed to change people's behaviour without relying on coercive measures (Thaler and Sunstein 2008). Today, behavioural interventions are applied within a wide range of fields (health, development, education, energy and environment, finance) and include a broad variety of tools (default rules, personalisation, disclosures of information, use of social norms) (cf. BIT 2017; SBST 2016; OECD 2017b). In this way, modern behavioural public policy goes well beyond the mere concept of nudge (Halpern 2016). Sunstein and Thaler's original approach, however, remains central to the debate on behavioural public policy, especially with regard to their normative justification for behavioural public policy, which is based on 'libertarian paternalism' (Sunstein and Thaler 2003; Thaler and Sunstein 2003). To provide justification, they advocate a framework of libertarian paternalism which preserves freedom of choice while steering people's behaviour to improve their personal welfare. Thereby, they propose a normative criterion of making individuals 'better off *as judged by themselves*'; that is, they argue that people's own preferences about their personal welfare should be taken as the relevant normative standard. To improve welfare thus means to satisfy individual preferences. In short, they rely on conventional welfare economics and the satisfaction of preferences as relevant normative criterion.

In this paper, I criticize Sunstein and Thaler's welfarist approach to libertarian paternalism as a normative framework for behavioural public policy. Instead, I develop a contractualist approach based on individuals as free and equal persons engaging in the participatory design of behavioural policies. I argue that based on normative deliberation and participatory collaboration, people may reasonably engage in collective self-commitment by designing choice architectures which nudge them into particular directions. In doing so, I provide a different normative justification for behavioural policy which does not run into the problems Sunstein and Thaler's welfarist approach is facing.

In short, I argue that taking a welfarist approach to the normative justification of behavioural public policy leads to methodological, epistemological, and practical objections. Given that people (often) lack stable and coherent preferences which is why they may fail to increase their welfare, preferences can no longer be simply identified with choices. In other words, behavioural economics challenges the validity of neoclassical axioms of rationality as adequate positive description of human behaviour. Eventually, this can lead to a reconciliation problem between behavioural and normative economics (McQuillin and Sugden 2012a). Based on this critique, I discuss two alternative approaches to behavioural public policy by R. Sugden and C. Schubert. Expanding on their concepts, I develop a contractualist approach which allows for a novel normative framework based on individual consent to collective choice rules, which are intended to align behaviour with values or achieve particular goals. By doing so, I establish a new participatory approach to behavioural policy, whilst also embracing the fundamental normative questions that are raised by behavioural economics.

I will proceed as follows. I will start by explaining why Sunstein and Thaler's welfarist approach fails, drawing on methodological, epistemological, and practical objections (2.).

Furthermore, I will argue that, even if their (implicit) assumptions were correct, they cannot show that their proposed methods actually increase individual welfare. I conclude by reflecting on the role of public acceptance and by discussing Sunstein's recent reply to similar objections. Next, I scrutinise the proposals by Sugden and Schubert as the only two remaining approaches that could reconcile normative and behavioural economics (3.1). Following this, I develop a contractualist framework based on participation and consent to collective choice rules (3.2). I discuss three possible objections to a contractualist approach (3.3). Finally, I conclude by pointing to the wider implications of the preceding discussion (4.).

In this paper, I avoid engaging in the broader debate surrounding nudges, behavioural interventions or libertarian paternalism and specific ethical implications (cf. Bovens 2009; Hausman and Welch 2010; Rebonato 2012; Hansen and Jespersen 2013; White 2013; Barton and Grüne-Yanoff 2015). Instead, I aim to escape this rather individualistic framing (Lepenies and Małecka 2015) and to focus greater attention on fundamental methodological and normative aspects of nudging (Sugden 2018; Infante et al., 2016a, b; Hausman 2016; Sunstein 2017c; Sugden 2017).

## The Normative Shortcomings of Libertarian Paternalism: Why a Welfarist Approach Fails

Libertarian Paternalism is the normative justification of behavioural public policy as provided by Sunstein and Thaler. It is based on conventional welfare economics, and thus defines welfare as the satisfaction of preferences. It aims to make people 'better off, *as judged by themselves*' (Thaler and Sunstein 2008, 5; italics in original). The normative criterion for assessing behavioural policy and to decide in which direction to nudge is hence revealed by the choices through which individuals aim to enhance their welfare. However, the findings of behavioural economics challenge the plausibility of revealed preference theory. Therefore, as Sunstein and Thaler themselves recognise, what matters are 'informed preferences'; the preferences people *would* reveal if they had complete information, unlimited cognitive abilities, and no lack of willpower (Sunstein and Thaler 2003, 1162). Due to cognitive limitations, people fail to fulfil their informed preferences, and make inferior decisions in terms of their own welfare. Although Sunstein and Thaler agree that actual revealed preferences may no longer provide a reasonable criterion of welfare, they still adhere to a normative criterion based on individual preference. That is why they assume that people's inconsistent choices may be treated as mistakes. They believe it is the context-dependence of choice that causes errors of reasoning and leads to poor decision-making. Choice architects thus need to reconstruct the informed preferences that people would act on if they were free from psychological biases – that is, limitations of attention, information, cognitive ability, or self-control. Sunstein and Thaler rely on an implicit welfarist model, taking the satisfaction of informed preferences as a normative criterion (Hédoin 2016; Qizilbash 2012; Sugden 2008). This view, however, leads to an approach of 'preference purification' (Sugden 2018; Infante et al. 2016a; Sugden 2015a; Lecouteux 2015a; Hausman 2012), 'laundered preferences' (Hausman et al. 2017; Hédoin 2016; Dold and Schubert 2016; Reiss 2013; Hausman 2012) and 'latent preferences' (Infante et al. 2016a, b; Fumagalli 2016; Sugden 2015a; Grill and

Scoccia 2015; Kahneman 1996), or 'informed desire' (Hédoin 2016; Qizilbash 2012; Sugden 2008), which raises serious objections. As a result, while dismissing the model of homo oeconomicus, Sunstein and Thaler appear to implicitly conceive of human agents as 'Faulty Econs', rather than as 'Humans' (Infante et al. 2016b).[1],[2] Distinguishing revealed and true preferences thus lies at the heart of Sunstein and Thaler's normative approach to justifying the implementation of nudges. In other words, their concept of libertarian paternalism assumes informed preferences and thus makes the critical distinction between revealed and true preferences the basis of their normative justification of nudges. This is why in order to critize their approach of normative justification the following section discusses the theoretical shortcomings of its theoretical underpinning.

A model of informed preferences must take a 'preference purification' approach. This approach treats human decision-making as consisting of psychological mechanisms that interfere with rational choice. Infante et al. (2016b) posit an 'inner rational agent' who is capable of generating consistent and context-independent preferences, but who is impeded and distracted by a psychological shell of ordinary human psychology. While the inner rational agent acts as a normative authority, preference purification aims to remove the psychological distortions and attempts to reconstruct the inner agent's true preferences. Sunstein and Thaler do not explicitly defend such a dualistic view. Yet they rely on informed preferences. Since these must be both inherently subjective and coherent, their approach requires some mode of reasoning that generates preferences that satisfy the conventional criteria of rational consistency. However, this preference purification approach lacks a psychological explanation of the process of latent reasoning, though it explains deviations from the process. This is problematic, at the least (cf. Kahneman 1996). Even if one accepts dual process theory – arguing that the inner rational agent represents System 2 and the psychological shell System 1 – one cannot simply assume that System 2 produces coherent preferences, especially given that, according to Kahnemann, System 2-processes are later add-ons and play a subordinate role (cf. Kahneman 2011). It seems implausible that System 2 processes could work separately to generate rational choices, not least because these processes may be evolutionarily more recent. Instead, decisions rely on both systems. Thus, there is no good reason to believe that informed, true, or latent preferences exist, or at least there seems to be no good reason to believe they do (Whitman and Rizzo 2015). This assumption runs afoul of the behavioural findings that libertarian paternalism is based on (Infante et al. 2016b), and seems to be psychologically ungrounded (Sugden 2015a). At one point, Infante, Lecouteux, and Sugden make the conjecture that this 'black box' approach to true preferences may originate from the structure of standard neoclassical theory. As rational choice theory is built upon axioms of consistency

[1] Accounts of informed or true preferences, as opposed to actual preferences, existed before the emergence of behavioural economics. See for example: John C. Harsanyi (1977). Similarly, concerns of endogenous or adaptive preferences have been raised elsewhere, see for example: A. Sen (1987, 45–47) or J. Elster (1983). However, as behavioural economics reveals the extent to which preferences depend on welfare irrelevant variables, informed or true preference accounts of welfare become particularly relevant for normative economics.

[2] Note that beyond libertarian paternalism there are different attempts to develop a notion of behavioural welfare economics (Bernheim 2016, 2009; Dold 2017). As they seek to integrate behavioural findings into neoclassical models, however, they remain within the conventional welfarist framework and face similar objections to those I raise against libertarian paternalism here. For some of the most advanced models of behavioural welfare economics, see for example Bernheim and Rangel (2007, 2009); Salant and Rubinstein (2008); Rubinstein and Salant (2012); Bordalo et al. (2013); Bleichrodt et al. (2001); Pinto-Prades and Abellan-Perpiñan (2012); Chetty et al. (2009); Kőszegi and Rabin (2007, 2008).

among preferences, completeness and transitivity, it does not aim to explain how an individual's preferences are constructed. However, as behavioural economics documents crucial deviations from these axioms, and as libertarian paternalism relies on preference purification to distinguish true from impaired and incorrect preferences, it seems implausible to adopt the same kind of instrumental 'neutrality' towards a mode of latent reasoning.

Moreover, as R. Sugden (2015a) shows, the concept of an inner rational agent is ungrounded, since no cognitive psychology theory or model is currently able to provide an empirical basis for it. Considering the isolation strategy of a preference purification approach, which needs to identify some component or mode of reasoning capable of generating context-independent preferences, Sugden discusses two specific models. One is from behavioural economics (Bordalo et al. 2013), and one is from decision field theory or decision-making under uncertainty (Busemeyer and Townsend 1993). While both models use an attention-based approach to evoke context-dependent choices, they both rely on the concept of latent preferences to exclude errors and analyse true preferences. However, in both cases the concepts of latent preference serve no explanatory purpose. Instead they merely stipulate a correct distribution of attention as one that conforms with predetermined correct preferences (Sugden 2015a). Clearly, this yields a corresponding definition of error as 'incorrect choice'. Yet, neither model can provide an explanation of which specific preferences are latent in the individual and which are not. Rather, they simply assert that there must be a mode of reasoning free from cognitive defects and weaknesses. Thus, the crucial distinction that latent preference are supposed to establish – rationality versus mistake – is unsubstantiated. In other words, what is missing is not only a description of the mode of reasoning that could generate context-independent preferences, but also a valid criterion to identify mistakes in reasoning. This leads Sugden to call the approach redundant and free-floating, even pre-scientific (Sugden 2015a, 586, 598). From another perspective, G. Gigerenzer (2015) substantiates the objection, arguing that Sunstein and Thaler's identification of behavioural errors is mistakenly based on narrow logical norms of rationality and, crucially, suffers itself from confirmation bias, as it only partially documents relevant research.

Finally, the model of an inner rational core appears doubtful, even according to dual process theory. In fact, the findings of behavioural economics have dismantled this dualistic and implicitly neoclassical model of human decision-making. It is thus somewhat ironic that, by implicitly relying on a preference-laundering view of welfare economics, Sunstein and Thaler seem to advocate the neoclassical model of homo oeconomicus as their normative role model, although or perhaps precisely because it has been renounced by behavioural economics as an inadequate positive description of human behaviour (cf. Dold and Schubert 2016; Schramme 2016; Schubert 2017; Whitman and Rizzo 2015; Angner 2015; Gigerenzer 2015; Berg and Gigerenzer 2010; Dold 2017; Rizzo 2017).[3]

[3] Infante et al. (2016b) argue that if behavioural welfare economics were to treat the purification of preferences as standardisation on a descriptive and pragmatic basis, rather than as identification of erroneous deviations, their views would not be open to such weighty objections. Behavioural welfare economics would then provide a model to standardise individuals' preferences to make them consistent with expected utility theory. Such an approach would need to proceed from a choice architect who makes the preferences compatible with rational choice theory according to the chooser's own judgements about her welfare. However, this is not what behavioural welfare economics understands as preference purification, especially not the version underlying libertarian paternalism. Beyond that, see E. Angner (2015) for a thorough discussion of the (epistemological) status of neoclassical theory within behavioural economics by comparison with Max Weber's notion of ideal types. Similarly, S. Heidl (2016) describes behavioural economics as de-idealisation of standard economic theory and says that it faces the same methodological limitations (cf. Lecouteux 2017).

Beyond this methodological argument, different objections may be raised against an informed preference view of welfare. One concern is that a true preference view is unrealistic and may overtax cognitively limited human beings (Qizilbash 2012; Sobel 1994). Qizilbash (2012) argues that Sunstein and Thaler rely on an implicit version of informed preference which leaves no space for human limitations. Elaborating on the information requirement of their welfare account, they appear to assume a third party, an ideal adviser for example (Railton 1986), who decides on behalf of the chooser to increase her individual welfare. However, given human limitations, an ideal adviser is implausible if one considers the information, knowledge, and capacities required to give the relevant advice. Yet, one may argue that only an approximation of welfare would be needed. If a third party could better approximate choices that increase an individual's welfare, libertarian paternalist interventions would indeed be possible. Thus, if one interprets Sunstein and Thaler as advocating a weaker version of the competent judge, their competent experts would only have to make better choices than the respective individuals, without needing to be perfectly rational or fully informed themselves. Crucially, however, such an approach seems to be prone to error (Qizilbash 2012). While I do think that they explicitly advocate a strong information requirement, this shows that their view of informed preferences may lead to serious implementation problems, since it either makes interventions impossible or provides flawed, error-prone guidance.

Furthermore, Sunstein and Thaler's normative approach also amounts to a more practical problem of knowledge or constrained epistemic access (Fumagalli 2016; Rizzo and Whitman 2009). Reconstructing people's true preferences would seem to require information and knowledge that might be unavailable to the choice architect. Furthermore, the information must be subjective in order to preserve the individualist notion of libertarian paternalism as expressed in the concept of 'better off *as judged by themselves*' (Grüne-Yanoff 2009, 2012). Sunstein and Thaler's approach therefore requires planners to have information not only about people's true preferences, but also about: their cognitive biases and limitations; how these biases influence agents' behaviour; the effect of interdependent biases; the choice contexts in which they manifest themselves; privately adopted self-debiasing measures and their effects; the effects of their interventions as well as the responsiveness of individuals' biases; and heterogeneity in the population with respect to these factors (Fumagalli 2016; Rizzo and Whitman 2009). A large part of the required knowledge may be inherently personal and local, depending on time and place, and as such in principle inaccessible to choice architects. Or the information may be tacit and difficult to communicate (Rizzo and Whitman 2009). But even if behavioural welfare economics could identify and reconstruct people's preferences, even an omniscient planner cannot gain knowledge of something that does not exist (Whitman and Rizzo 2015). While one may conclude that this knowledge problem prevents interventions with insufficient information, it also makes a case against solely relying on information from laboratory experiments, and indicates the problematic implications of such interventions for real political processes. Insofar as welfare economics addresses imagined policy-makers, the benevolent planner of libertarian paternalism faces constraints and challenges of real-world democracy, public choice and the political economy, and might indeed be subject to biases him- or herself (Banuri et al. 2017; Hirshleifer and Teoh 2017; Schubert 2017; Schnellenbach and Schubert 2015; Pasche 2014; Sugden 2013; Glaeser 2006). Besides requiring perfectly informed, impartial and benevolent political actors, Rizzo and Whitman (2009) believe that choice architects are likely to rely on rules of thumb or appeal to their own preferences or those

of (self-appointed) experts. In this way, libertarian paternalism may tend to reinforce socially approved preferences and provide a rather conservative concept which strengthens status-quo norms (Schnellenbach 2012).[4]

Finally, when adopting a preference purification approach, libertarian paternalism requires eliciting people's true preferences, which cannot simply be observed in their actual choices. In response to this challenge, Sunstein and Thaler propose a version of cost-benefit analysis to measure the welfare effects of different designs, and some rules of thumb that could provide proxies for welfare, in situations when direct analysis is not possible (Sunstein and Thaler, 2003, 2006; Thaler and Sunstein 2003). Although this aspect is crucial to their approach, Sunstein and Thaler only touch upon this issue very briefly. Suggesting a version of cost-benefit analysis that cannot be based on people's willingness to pay, but instead must be more open-ended and even subjective, they end up discussing the gains and losses of default rules and automatic enrolment in different programs. Although they acknowledge how difficult and expensive such an endeavour will be, they defend cost-benefit analysis as a method for evaluating welfare effects. Yet it remains unclear how a discussion of gains and losses relates to individual preferences and thus welfare. Why should one assume that an evaluation of gains and losses has anything to do with individual preferences and that it can define welfare? They seem to believe that a cost-benefit analysis that reveals outweighing benefits must somehow represent an individual's preference. They simply assume without justification that this is plausible (Sugden 2008). It is therefore striking that the first method they recommend for assessing true preferences ends up in a discussion of trade-offs, without even attempting to appeal to preferences and welfare. Admitting the complexities of such an approach, they suggest three rules of thumb for use when a cost-benefit analysis is impracticable. First, they suggest minimising the number of opt-outs, since this leaves more people sufficiently satisfied. A second rule of thumb they suggest is to follow the direction the majority would choose. Lastly, they recommend forcing individuals to make their choices explicit. In all cases, however, it remains unclear how they would identify and promote individual welfare. Their approach again ignores the question at stake; namely, those preferences that might be context-dependent and inconsistent. Whether minimising opt-outs or following the majority choice, the choices of the consistent and rational choosers will define the relevant preference while the preferences of inconsistent choosers will not contribute to the final outcome (Goldin 2015). In other words, Sunstein and Thaler's proposed methods for assessing true preferences fail to prove that choice architects will succeed in increasing individual welfare. Instead, their rules will increase what planners *believe* to be an individual's welfare. Therefore, Sunstein and Thaler's vagueness concerning how to elicit people's true preferences appears to be genuinely conceptual and not merely practical, as they have suggested (Sugden 2008).

[4] Another objection is that determining what it means for an individual to have complete information, unlimited cognitive abilities, and no lack of self-control, is itself a difficult task. It is also inescapably normative, as it involves substantive assumptions about welfare (Fumagalli 2016; Sugden 2009). To put it simply, it may require critical normative hypotheses to infer what, say, complete information might imply for an individual and her preferences.

This leads to another question about the relation of a purified state of complete information, and welfare (Hausman 2016). Even if it were possible to determine what an individual would choose if they had complete information and no reasoning impairments, this choice is not necessarily identical to their welfare – at least there is no guarantee that this is true. As T. Cowen (1993) puts it, perfectly informed preferences might not always be relevant for actual, imperfectly informed choices in the real world. Knowing what an individual would want if endowed with perfect information doesn't necessarily provide helpful information about what would increase an individual's welfare now.

Whether or not this may induce an approximation view of preference (Qizilbash 2012; Hausman 2012), Sunstein and Thaler's welfare criterion fails to deliver the promised normative foundation justifying behavioural interventions. Sunstein and Thaler provide support for their view with anecdotal examples, to which most of their readers are likely to agree. They even seem to argue that, at least in some cases, people may be willing and grateful to be nudged (Thaler and Sunstein 2008, 107). Yet, as they themselves insist, the only valid criterion is individuals' own judgements. It is in this context that one may query the status and role of empirical surveys about whether or not people 'like' or 'want' behavioural policy tools such as nudges (cf. Sunstein 2017a; Jung and Mellers 2016; Hagman et al. 2015). Political relevance notwithstanding, one must be careful not to blur the line between public acceptance of policy tools, and their justification and fundamental legitimacy. Because the normative framework of a welfarist libertarian paternalism is methodologically flawed and cannot show that it effectively increases individual welfare, the shift in Sunstein's recent research towards an empirical approach (collecting public opinion on nudges) may appear to be a new different justificatory strategy (cf. Sunstein 2017a, b, 2016; Sunstein et al. 2017; Reisch et al. 2017; Reisch and Sunstein 2016). Sunstein's move may therefore indicate the normative problems of a welfarist underpinning of libertarian paternalism. However, surveys cannot address ethical issues around acceptance, nor account for the complex issues and mechanisms of behavioural interventions.[5] So investigating the social acceptance of particular nudges cannot compensate for the missing normative framework of behavioural public policy.

Sunstein (2017c) recently addressed similar objections by responding to a paper from R. Sugden (2016). In his response, Sunstein treats the objections to his normative criterion as empirical questions rather than profound methodological concerns. He claims that by distinguishing four different categories of cases in which nudges can be applied, he might be able to substantiate the normative criterion of 'better off *as judged by themselves*'. Sunstein thereby implicitly approves of a true preference account of welfare. However, while there are obvious cases in which the criterion provides reasonable guidance, Sunstein acknowledges the difficulties when people lack latent or antecedent preferences and preferences are not consistent.[6] Yet it is precisely in these cases that behavioural economics and behavioural public policy is interested. Therefore, as Sunstein limits his normative approach to the restriction of possible solutions which may help choice architects to *orient* themselves, he seems to recognise the imperfection of a normative framework of a welfarist libertarian paternalism. But again, he fails to realise the implications of assuming an inner rational agent. By treating the issue as empirical rather than as fundamentally normative and conceptual, he is unable to conceive the full scope of the problem (cf. Sugden 2017). Instead, Sunstein reinforces his (new) empirical stance and appeals to public acceptance of behavioural interventions. Thaler

[5] Additionally, one may ask what behavioural findings may imply for methods of surveys and questionnaires themselves given such mechanisms as framing, anchoring, or availability effects (cf. Sunstein 2017a). Moreover, see also Tannenbaum et al. (2017) who show that there might be a 'partisan nudge bias' in people's evaluation of nudges (see also Fox and Tannenbaum (2015)).

[6] As Sunstein (2017c) emphasises, there may be clear cases where one might reasonably assume latent preferences. Beyond Sunstein's examples of apparent antecedent preferences and cases of self-control problems (cf. Sugden 2016), one might also imagine complex tariffs which leave individuals worse off as they do not choose the cheapest alternative (Infante et al. 2016b). However, such an identification of subjective (alternative tariffs) and objective preference rankings (cheapest prices) is not always possible. In fact in many cases there may not even exist such an obvious analogue as objective ranking. This highlights the proximity of a preference purification view to objective approaches of welfare, in sharp contrast to Sunstein and Thaler's emphatic claim of a neutral and formal kind of mere means-paternalism.

and Sunstein imply in *Nudge* (2008), that people's gratitude for behavioural interventions serves as a justification, and such interventions are 'warmly welcomed' (Sunstein 2017c). However, I argue that the lack of a valid normative framework cannot be compensated for by appealing to positive public attitudes towards nudging. Ultimately, Sunstein's response does not provide an answer to the methodological and normative critiques.

Eventually, advocating an informed, purified, or true preference view of welfare may be seen as a move towards a more objective account of welfare and towards a notion of what *should* be preferred rather than what people do in fact prefer (cf. Hausman 2012; Scanlon 1998; Railton 1986; Griffin 1986).[7] In this way, libertarian paternalism may take the middle ground by providing a normative framework for behavioural public policy. Crucially, though, Sunstein and Thaler do not take a stance on welfare, but aim to rely on the formal, instrumental, and neutral welfare criterion of revealed preference. Yet, as we have seen, in the context of behavioural sciences and economics this amounts to an informed preference or preference purification approach, which faces severe problems. Avoiding a substantial normative position therefore fails as a justificatory strategy. While they wish to evade a substantial normative argument by building their normative framework of libertarian paternalism on a recourse to conventional welfare economics, they miss the complexities of welfarist assumptions within a behavioural context. As a result, I argue that behavioural and normative economics cannot be reconciled based on the assumption of true preferences, and the reconciliation problem persists for the justification of behavioural public policy (McQuillin and Sugden 2012a).[8]

[7] Consider one of Sunstein and Thaler's most well-known examples. A cafeteria director must choose how different food items are presented (the opening example in *Nudge* (Thaler and Sunstein 2008). She notices that some customers are inclined to choose those items that are more visibly displayed. Sunstein and Thaler believe she should try to arrange the items such that she highlights those that customers would choose themselves, since this would best satisfy the criterion of 'better off *as judged by themselves*'. However, people do not always have stable and context-independent preferences, and the arrangement of items has a significant effect on people's preferences. In other words, customers' well-ordered, true preferences may not formally exist (Sunstein and Thaler 2003; Thaler and Sunstein 2008). Therefore, the strategy of choosing what the customer would choose on his own collapses as it is not possible to identify his preferences independently from the director's arrangement (Sugden 2008). Sunstein and Thaler suggest the only remaining strategy is to make the customers best off, all things considered. While this is clearly paternalistic on their own terms, it preserves freedom of choice insofar customers may still make their own (unhealthy) choices if they wish. However, the normative criterion that guides the director is straightforwardly paternalistic. The director is supposed to arrange the items as to what *she thinks* would make the customers best off, all things considered. Sunstein and Thaler's seem to appeal to some self-evident objective value of 'healthy behaviour' which suggests that the director's assessment cannot be different from the customer's true preferences. This, however, establishes a substantive normative assumption which would need an explicit and rigorous justification. Nevertheless, by referring to obvious examples that no one would reasonably object to, Sunstein and Thaler may seem to overcome the conceptual normative fallacy by implicitly invoking an objective account of welfare based on allegedly common values (cf. Whitman and Rizzo 2015; Sugden 2016). I do not claim that the paternalistic stance of libertarian paternalism is problematic in itself, but it should be explicitly defended in order to provide a valid normative framework.

[8] Prior to McQuillin and Sugden's (2012a) image of a reconciliation problem, N. Berg (2003) argued for a notion of normative behavioural economics. As this approach does not specifically address libertarian paternalism, it may be neglected here. However, there seems to be no contradiction between the reconciliation perspective which may rather be perceived as behavioural normative economics (Dold and Schubert 2016) and a notion of normative behavioural economics. Instead, it demonstrates the importance of the links between behavioural and normative economics on a theoretical, conceptual, and methodological level. D. R. Just (2017) recently introduced the 'behavioural welfare paradox', which addresses the same issue, namely the paradoxical consequences of behavioural economics for standard welfare analysis. Similarly, T. Grüne-Yanoff (2009) described the tension between welfare economics and behavioural economics' findings within libertarian paternalism as a 'soft paternalist's paradox'.

## From (True) Preferences to Normative Agreement

### Salvaging Normativity: R. Sugden, C. Schubert, and beyond

To reconcile behavioural and normative economics, Robert Sugden has advocated a concept of 'freedom as opportunity' as normative criterion (Sugden 2018; Schubert, 2015b; Sugden 2010, 2008, 2007, 2006, 2004; McQuillin and Sugden 2012a, b). Based on the insight that conventional welfare economics' normative criterion of preference satisfaction fails in light of behavioural findings, Sugden's approach integrates incoherent preferences into the broader liberal concept of consumer sovereignty (Sugden 2004; McQuillin and Sugden 2012a). He contends that regardless of whether people reveal coherent preferences, they value opportunities. That is, people value the freedom to choose whatever preference they want act on. It is not the satisfaction of specific preferences which provides the normative criterion, but rather the opportunity to act on individual preferences, whether they are coherent or not. Even if they cannot be rationalised by a single set of preferences, an individual may then identify with all of her choices. Following a continuing-agent view as opposed to a conventional multiple-selves approach of time-inconsistent behaviour, she may be represented as a responsible rather than a rational agent (Sugden 2015b, 2007; 2004). In this way, Sugden offers a contractarian approach, arguing that a normative analysis depends on each individual's subjective understanding of value (Sugden 2006). While such a contractarian perspective considers each individual's subjective welfare, it is argued that mutual benefit is best achieved through the market (Sugden 2018, 2008, 2004). Thus, even though people may not have coherent preferences, the market provides the best and most efficient mechanisms for providing people with opportunities. Sugden therefore argues that normative and behavioural economics can be reconciled with a contractarian and market-based approach, without relying on any form of paternalism.

Christian Schubert, by contrast, has argued for 'preference learning' as a dynamic alternative concept of opportunity (Schubert 2015b; Schubert and Cordes 2013; Dold and Schubert 2016; but also: Schubert 2015a, 2012a, 2012b). Defining learning as the voluntary and cumulative acquisition of new preferences, the claim is that individuals value the opportunity to learn and want to maintain the set of potential preferences that they have the capacity to learn, if they choose to do so (Schubert 2015b). The normative criterion of welfare thus consists in people's ability to engage in the learning of new preferences (Schubert and Cordes 2013). However, the opportunity to learn includes the possibility of self-constraint. As people value their individual sets of opportunities to learn, they may wish to influence how they develop over time. To maintain their opportunities to learn, people may voluntarily choose devices of self-commitment or even self-constraint in order to not endanger their own learning dynamics (Schubert 2015b). Though following Sugden in taking a contractarian perspective, Schubert does not think that people prefer to maximise opportunities, but might instead be able to handle only a limited set of opportunities, which is why optimising their opportunities to learn new preferences may involve legitimate instruments of self-constraint.

Independently of whether people's preferences are coherent or rational, Sugden argues that people prefer increases in opportunity, which they seek to maximise. Sugden's responsible agent acknowledges responsibility for all past, present, and future choices. Yet his approach does not allow for attitudes towards future preferences other than unconditional endorsement

(Schubert 2015b). But since the mere increase of options may not lead to an increase in welfare, and since agents may become overloaded by too many choices and opportunities, Schubert proposes optimisation of opportunities *over time*. To this end, individuals might appeal to measures of self-constraint to optimise their future set of opportunities. Differentiating self-command and self-constraint, Schubert therefore defends his opportunity to learn criterion as a dynamic variant of Sugden's opportunity criterion. It enables people to engage in self-commitment in order to prevent choice overload from reducing future opportunities to learn (Schubert 2015b). Additionally, Schubert (2015b) writes that Sugden's approach seems to favour the impulsive over the reflective self and, furthermore, that people might want to constrain themselves for considerations of subjective coherence. Yet, as Sugden emphasises, the responsible agent's unconditional endorsement merely delegates future decisions to one's future self. In addition, people's preferences for self-constraint might be much less common than Schubert suggests (Sugden 2015b). To Sugden, it is somewhat paradoxical that a theory valuing opportunity could involve opportunities to reduce future opportunities. Based on the principle of consumer sovereignty, Sugden claims that his account may even include demands of self-constraint, even though it would not be chosen by the idealised responsible agent. However, as the market is justified by its power to provide, satisfy, and even create opportunities, it hardly seems compatible with preferences for self-constraint, since the market allows everyone to get whatever she wants and is willing to pay for, *when she wants it and is willing to pay for it* (Sugden 2015b). While Sugden aims to preserve the neutrality of a formal account of normativity, Schubert queries whether individuals wish to maximise their opportunity sets. He argues that people value opportunities to learn new preferences, rather than to satisfy whatever preferences they happen to have.

Although I would not call Schubert's view perfectionist in the sense of assuming an objective account of the good based on its intrinsic value (cf. Sugden 2015b), by defining the formation of new preferences through learning as a normative criterion, it moves in a more substantive direction. I agree with Sugden that the opportunity to learn seems to be a second-order concern, and cannot establish a general normative criterion, unlike the neutral and liberal account of opportunity based on consumer sovereignty. However, I follow Schubert in dismissing Sugden's approach as too neutral and market-biased. A purely individualistic, market-based approach does not sufficiently take into account agents' differentiated and nuanced attitudes towards their own (future) preferences as highlighted by behavioural findings, which is why individuals should be able to engage in some form of self-commitment and self-constraint. While Sugden intends to integrate cases where people fail to reveal coherent and stable preferences, it seems implausible to me to eschew any qualification of preferences in favour of the simple maximisation of opportunities whatever they happen to imply. Whether or not people have such a passion for increasing their opportunities, agents should have some means by which to normatively shape their preferences. I agree with Schubert that people care about optimisation of opportunities over time, which requires self-commitment and self-constraint. Even though markets may provide the most efficient mechanisms for maximising individuals' opportunities, there may be other ways to retain the liberal principle of consumer sovereignty while decoupling it from the orthodox assumption that people act on coherent preferences. Therefore, while I join Sugden and Schubert in adopting a contractarian path, I will elaborate further and try to develop a contractualist perspective to reconcile normative and behavioural economics.

## Developing a Contractualist Approach to Behavioural Public Policy

Sugden's and Schubert's contractarian approaches are based on the assumption that each individual seeks to advance his self-interest, and that cooperative behaviour creates mutual benefits. By way of fair agreements, individuals engage in mutually beneficial transactions which are facilitated by the market and its mechanisms of cooperation. A contractarian perspective thus takes the individual and her personal and subjective interests as starting point, and aims at mutually advantageous outcomes by maximising individual interests. In so doing, it serves as normative foundation for social cooperation as facilitated by the market. This notion of contractarianism stems from Thomas Hobbes and has been vindicated in modern times by David Gauthier (1986), but also by James M. Buchanan (1975). It has subsequently been considered as a basis for ethical theory and economic ethics (Luetge 2005; Luetge et al. 2015).

By way of contrast, a contractualist approach has its roots in Jean-Jacques Rousseau's work, and in a social contract tradition that is more interested in the conditions that can justify the pursuit of one's interests to others.[9] Based on the equal moral status of persons, contractualism takes a more egalitarian view and aims at agreements for social cooperation that individuals would agree to, from a perspective that respects their equal moral status as rational autonomous agents (Ashford and Mulgan 2012). Contractualist approaches are thus grounded on individuals' agreement on the principles that govern social cooperation and individual behaviour on an egalitarian basis. T.M. Scanlon (1998) has described the contractualist criterion as principles nobody could reasonably reject. Most notably, John Rawls (1971) adopted a contractualist approach to derive his principles of justice. By proposing a veil of ignorance behind which individuals would not know their real, current social status, Rawls argues that people would use the 'maximin' strategy to deduce the difference principle as a principle to which everyone would agree. Narrow contractualism requires no veil of ignorance; rather than principles people *would* agree to, Scanlon is concerned with principles nobody *could* reasonably reject. Be that as it may, I take Rawls' account as providing a two-tier system and advocating a liberal political order that includes the possibility of a contractualist justification of (soft) paternalist interventions (Rawls 1971; Herzog 2008; Ferey 2011; Hédoin 2016).

Similar to G. Dworkin's understanding of paternalism (Dworkin 1972, 2017), Rawls proposes a contractualist approach to paternalism justified by individuals' own acknowledgement 'in the original position to protect themselves against the weakness and infirmities of their reason and will in society' (1971, 249–250). While Rawls defends interventions based on substantive welfare and primary goods, he is equally committed to the liberal principles of pluralism, subjective interests, and consumer sovereignty, so that his contractualist notion of paternalist interventions may even be compatible with an informed preference approach to welfare (Hédoin 2016). Reasonable persons may therefore agree over a paternalistic collective choice rule, if they acknowledge that it is in their subjective interest to do so. Following A. Sen's distinction of preferences and values (1970, chap. 5, esp. pp. 64–67), people *as persons* may agree to paternalist interventions if they recognise that they may not always reveal

[9] Compare J. Heath's distinction of micro- and macrocontractualism in which he defines contractarianism as a distinct kind of microcontractualism (Heath 2014, 145ff.). See also Hausman et al. (2017, 224ff.) who despite small differences distinguish between perspectives seeking mutual advantage and those which focus rather on impartiality or reciprocity.

coherent preferences. But as continuing agents they have interests in superior values, making it reasonable to agree to a collective choice rule (Hédoin 2016).[10] In order for such paternalist interventions to be justifiable, a robust notion of consent is needed to establish the foundation for collective choice rules (Dworkin 1972). As G. Dworkin shows in a thorough discussion of J.S. Mill's account of paternalism, it might be reasonable for rational individuals to agree to institutional arrangements as 'social insurance policies' that protect them from the irrational decisions that they might make in certain temporary states (Dworkin 1972, 78–81). Or, as he puts it: 'What I am looking for are certain kinds of conditions which make it plausible to suppose that rational men could reach agreement to limit their liberty even when other men's interests are not affected' (Dworkin 1972, 78). Taking the classical example of Ulysses, who asks his crew to restrain him so that he might hear the Sirens' deluding song without acting on it, Dworkin argues that there must be 'genuine consent and agreement' in order to justify such paternalist interventions (Dworkin 1972, 77).[11] What I propose is thus a contractualist perspective justifying paternalist interventions as collective choice rules based on consent. It is plausible that an individual, recognising her behavioural biases and cognitive impairments, could consent to a collective choice rule based on her (superior) values in order to protect herself from unreasonable, temporary desires and biases, and to preserve personal integrity.

With an approach based on collective choice rules and consent, behavioural interventions can be conceptualised as opportunities for collective self-commitment. Along these lines, L. Heidbrink suggests that questions of nudging and libertarian paternalism be reframed in terms of sustainable and political self-binding (Heidbrink 2015; Heidbrink and Reidel 2011; for a general introduction to the issue of rational self-binding, see also Schaal 2009; and especially: J. Elster 1979, 36–111). He claims that nudges are justifiable as forms of self-binding if they are transparently designed and people consent not only to their implementation but also to their welfare objectives. Individuals may thereby accept limitations to their freedom while protecting individual and collective autonomy as defined by the personal and political self-conception of a society. In this light, the notion of a contractualist framework based on collective choice rules relates to concepts of individual self-binding which advocate the voluntary limitation of one's freedom in order to realise self-chosen goals (Heidbrink 2015, 186). One can distinguish at least two different types of self-responsible self-binding (Elster 1979, 103–105); the exogenous manipulation of an actor's environment, and the endogenous manipulation of the actor's character. Self-binding procedures through the limitation of external and internal options may be delegated to institutions (Heidbrink 2015, 187; Herzog 2008, 120–121), such as collective choice rules. But for such institutions to be legitimate, individuals must not only be the authors of self-binding procedures, but they also must be actively involved in their development. The institution of deliberative self-binding procedures thus requires the democratic participation of citizens (Heidbrink 2015, 184–185). The

[10] The distinction of values as attached to persons and preferences as ascribed to selves draws on K. Arrow (1963) who distinguishes interests from values, but it relates particularly to J. Harsanyi (1955) who distinguishes an individual's subjective preferences from her ethical preferences which are those had she an equal chance of being in anyone's position (cf. A. K. Sen 1970, 66).

[11] Note that while G. Dworkin (1972) rather vaguely refers to some certain kinds of temporary conditions and humans' limited cognitive capacities due to which individuals tend to make irrational decisions, today behavioural findings are able to provide reliable evidence on how human choices are shaped and depending on normatively irrelevant factors. Interestingly, however, he points to cooling off-periods which have become an integral part of Sunstein and Thaler's concept of nudge (cf. 2003) and of behavioural public policy in general (cf. Lynch and Zauberman 2006).

discussion of self-binding therefore reveals the importance developing a notion of consent beyond mere agreement, which is based on participation and deliberation. Accordingly, the contractualist framework involves a substantive notion of consent since it calls for the active participation and deliberation of all those affected by collective choice rules, and demands the engagement of people as citizens. Genuine agreement to behavioural intervention in the form of a collective choice rule is only possible when individuals actively deliberate and participate in the decision about its aim and purpose, its effects and consequences, and its form and substance. Thus, the contractualist approach that I propose depends on a collaborative way of developing interventions and policies, and requires citizens' participation in processes otherwise limited to policy-makers, economists, psychologists, and other experts.

Relying on a participatory approach may not only increase knowledge for policy-makers and (better) encourage behaviour change (John 2013; John et al. 2011), but can also provide a normative justification for behavioural public policy by applying a responsive and citizen-based approach beyond technocratic limits. In so doing, a participatory contractualist framework institutionalised in collective choice rules advocates a (more) deliberative economy (Anand and Gray 2009).[12] The recourse to deliberative strategies may also strengthen the justification of behavioural policies from a legal perspective, by allowing individuals to learn and eventually collectively agree upon normative aims (cf. van Aaken 2006; Schaal and Ritzi 2009). As van Aaken (2015) demonstrates for the field of administrative law – which, by the way, may prove to be particularly relevant for the implementation of nudges –, the application of deliberative methods may be used to advance consent and approval based on behavioural findings, as, for example, in terms of behaviour driven by fairness considerations or regarding aspects of procedural fairness. Thereby, people may bind themselves to protect their 'deeper' values against their cognitive weaknesses and their irrationality (Elster 1979, 84, 111). In accordance with anti-perfectionist liberalism meaning that governments and policy-makers should not support one particular and objective idea of the good based on its intrinsic value (Moles 2015), individuals as free and equal cititzens may engage in the development of behavioural interventions such as nudges and design choice architectures in order to enhance their welfare or pursue other normative goals they have agreed upon, such as liberty (McPherson and Smith 2008), equality (Smith and McPherson 2009), fairness (Hacker 2016a, b), or other deontological criteria (Lecouteux 2015a, b).[13]

In response to the question of who the addressee of normative economics is supposed to be (cf. Sugden 2013), a contractualist approach focuses on individuals and citizens, who in a deliberative economy officiate as planners, choice architects, and policy-makers. While a welfarist approach assumes policy-makers or economists are able to decide on the normative framework in terms of individual welfare, I deem it necessary that individuals themselves may deliberate and co-decide on the values and normative principles on which policies and choice

[12] See also Lepenies and Małecka (2016; 2015) who focus on the institutional implications of behavioural policies and argue that nudges require different legal measures as institutional safeguards against their possible negative consequences. They explicitly take an institutional rather than individualistic perspective. I adopt their institutional perspective, but aim at the normative justification provided by the framework of libertarian paternalism. Thus, I take my contractualist approach as institutional proposition while being based on (the institutionalisation of) collective choice rules.

[13] The assumption is that conventional welfare economics takes a preference-based utilitarian stance, and a contractualist framework may eventually overcome the utilitarian bias in economics by enhancing rather deontological aspects.

architectures are designed.[14] It is precisely because individuals do not always act on coherent and stable preferences that they may wish to engage in some form of self-binding as collective choice rule and agree on a choice architecture and an according normative frame and objective. The contractualist perspective may therefore provide a bottom-up approach to nudging since it seeks to generate nudges on a deliberative and participatory basis rather than by a top-down definition of policies and their normative purposes (cf. Moseley and Stoker 2013). So in contrast to Sunstein and Thaler's implicit welfarist foundation of libertarian paternalism, I propose a political justification whereby behavioural policies are grounded in a political process of normative deliberation and collaborative participation (cf. Guala and Mittone 2015).[15] In doing so, I agree with arguments that nudges can be legitimate insofar they are part of democratic processes (cf. Nys and Engelen 2017). Yet, my proposal goes even beyond this, since it does not simply refer to existing political procedures of democratic systems, but requires greater participation and involvement of the individuals concerned. In short, this is necessary due to the specific behavioral and at least partially hidden nature of nudges or other instruments. Based on a justification of nudges as forms of collective self-binding, I argue that behavioral policies in order to be legitimate must be agreed and consented to by all individuals affected. It is this form of genuine agreement and consent that requires peoples' active participation in the process of developing behavioural policies. Admittedly, compared to their welfarist approach, a contractualist normative perspective may well limit the scope of possible behavioural interventions. However, this indicates the complex normative deficiencies of a welfarist view, rather than a flaw in the contractualist framework.

In the above argument, I established a contractualist approach for the normative justification of behavioural public policy. Distinguishing preferences that vary over time from values that are persistent and deeply held, I have advocated a contractualist approach in which individuals consent to collective choice rules through normative deliberation and participatory collaboration. Agreeing on a normative framework and its guiding principles and aims, people may reasonably engage in collective self-commitment by designing choice architectures which nudge them into particular directions.[16] Individuals as persons may choose to design choice architectures in order to counter-nudge their behavioural biases in compliance with explicit normative goals and values (Sibony and Alemanno 2015; Baldwin 2014; Schubert 2014). I thereby agree with Hacker (2016a) that the contractualist perspective moves beyond the

[14] Sugden argues that based on a particular understanding of normative economics, a welfarist approach characterises politics as executive action, while his contractarian model describes politics rather as negotiation to achieve mutually beneficial outcomes (Sugden 2013, 529). A contractualist perspective, I might add, emphasises the deliberative character of the complex reality of politics to achieve fair agreements between individuals as autonomous moral equals.

[15] Guala and Mittone (2015) similarly suggest a political, though contractarian, justification of nudges. However, their approach relies on negative externalities which biases of human decisions may cause and which they argue justify the employment of behavioural policies.

[16] Schnellenbach (2011) points at two empirical case studies of individual self-binding. In the first, employees voluntarily chose a contract option which gave them less money if they failed to achieve a certain productivity goal which they agreed on in advance (Kaur et al. 2010). The second concerns retirement savings based on Benartzi and Thaler's popular experiments and their *SMarT* plan (Thaler and Benartzi 2004; DellaVigna 2009). Another example is provided by *sticK*, a company founded by Yale economist Dean Karlan which enables its customers to engage in individual commitment contracts (The Economist 2008). Additionally, there is one empirical study on individuals' motivations for engaging in behavioural policies. It finds that paternalism is not demanded by people who need it as a commitment device (Pedersen et al. 2014) which may provide an argument for paternalism rather as a device for collective self-commitment (Schubert 2017).

neoclassical-behavioural dichotomy and openly embraces the fundamental normative questions at the heart of behavioural public policy. The contractualist approach expands the contractarian endeavours, and by presenting a new avenue for the normative justification of behavioural public policy, it may eventually reconcile normative and behavioural economics.

## Three Possible Objections

In the following section, I will briefly discuss three possible objections to my contractualist proposal. The first concerns the overload of individuals, the second raises the question of whether a contractualist approach faces the same problems as a true preference-account, and the third asks whether a contractualist perspective still counts as paternalism at all.

One objection to my contractualist approach is that, it burdens individuals with too much responsibility and information, and that individuals are generally not the appropriate addressees to discharge these responsibilities. For example, as behavioural insights are increasingly applied in the field of consumer policy (see, for example: OECD 2010, 2014, 2017a; Vringer et al. 2015), one might argue that the state and the government should intervene and regulate markets in line with laws, principles, norms, and (societal) goals, rather than individuals and consumers themselves. Touching on consumer responsibility, two responses are possible. First, it is an objection not just to the contractualist approach but to behavioural public policy in general, since it is a fundamental feature of this approach that it applies to individuals' choices and decision contexts. However, as a welfarist underpinning fails to provide sufficient justification and the contractualist approach aims to fill this justificatory gap, I do not address the particular question of when or whether behavioural policies should be applied at all. My claim is that *only if* behavioural policies should be applied, they can be justified by a contractualist approach rather than a welfarist framework. My second response targets the institutional backdrop of my contractualist proposal. I advocate a contractualist approach based on collective choice rules to which individuals voluntarily consent through deliberation and participation. It is the state's responsibility to enable collaboration throughout the policy making process. The contractualist approach explicitly requires the state to establish an adequate structure and institutional framework which facilitates deliberation and participation in the design of collective choice rules. There exists some disagreement about whether behavioural public policy should be seen as a supplement to traditional policy tools or whether it should replace conventional measures, such as bans, mandates, or incentives. I deem it reasonable for behavioural approaches to be employed if there is sufficient evidence that they significantly improve the status quo of regulations and policies. Thus, I do not suggest delegating new responsibilities to individuals that have previously been fulfilled by the government. Instead people's decision-making processes should only become the subject of public policy if behavioural findings provide solid and reliable ground for substantial improvements. Crucially, though, it is precisely the participatory approach of the contractualist framework that empowers individuals to decide themselves whether or not they want to engage in particular collective choice rules. Therefore, rather than being charged by the government with too much responsibility, people are enabled by the contractualist approach to make

use of behavioural findings to increase their individual welfare or achieve particular outcomes. Finally, digital technologies can make citizen participation easier (cf. John 2013) and thus help realizing such bottom-up approaches in policy-making. Recognizing the risks and negative implications of digital technologies which shift nudging to a whole new level using for example Big Data and individual targeting on Social Media platforms, digital technologies and the Internet still have a (radical) democratic potential. Think of tools, apps and open source-activities which allow, among other things, facilitating the organisation of communities (from visualization and mobilization to personal consultation), joint discussions and the negotiation of interests, including voting. In short, the participatory and deliberative approach that I call for can be supported and facilitated through new (communication) technologies which harness its democratic and empowering potential in order to design collective choice rules. Today, there is already a number of groups working on utilizing digital technologies for more participation and democracy, engaging in broader movements of technology for democracy and for the wider (social) good. Think of, for example, open government initiatives, social entrepreneurs, co-creation methods or 'civic tech'.

A second objection is that the consent-based contractualist approach must assume that individuals are able to agree upon a collective choice rule according to their true, rational preferences, and the approach therefore faces the same difficulties as an informed preference view (cf. Binder 2014). However, as theories of deliberation and deliberative democracy show, people can learn new preferences, interests, and perspectives through deliberative practices, and can thereby make epistemic improvements (Schaal and Ritzi 2009, 62–69). I do not argue that people will achieve 'full' rationality or act exclusively on their second-order preferences. However, individuals can reflect on their preferences and choices through deliberative and participatory procedures, and thus they can agree on choice architectures which respond to behavioural anomalies, without presupposing the existence of informed, true, or 'rational' preferences. The contractualist approach does not assume fully rational individuals or an inner rational agent, but instead assumes that reasonable individuals may, as persons, agree on collective choice rules according to specific normative goals and values which they tend to ignore or contravene in concrete decision contexts. For example, in an experiment with policy-makers of the World Bank and the Department for International Development in the United Kingdom, Banuri et al. (2017) show that group deliberation may reduce sunk cost bias and confirmation bias. However, deliberation may have no effect on some biases, or may even exacerbate them. Thus, deliberation may be more effective for counteracting some cognitive impairments but not others (Banuri et al. 2017, 26). Even though these results are not easily generalised, a deliberative and participatory approach to collective choice rules is not necessarily a panacea. Yet, even if it is not equally effective for all behavioural impairments, various formats and designs might be developed in order to align effects with desired outcomes. In any case, it is reasonable to assume that people can reflect on their choices in deliberative processes and realise at least some of their psychological and cognitive constraints, without assuming an inner rational agent, true preferences or 'full' rationality. Therefore, the contractualist framework does not face the same objections as an informed preference view.

Lastly, a third objection is that by reframing behavioural interventions as collective choice rules based on consent, the contractualist approach is no longer paternalism as defined by G. Dworkin (2017) (cf. Binder 2014; Hausman and Welch 2010). Paternalism requires the intervention to be against the will of the individual, while the contractualist approach is based

on an individual's consent. The point, however, is that 'will' cannot be defined as identical to preferences. As shown by behavioural economics, people sometimes fail to act on their preferences, which is why one may say that people sometimes do not want what they *want*. Agreeing on a collective choice rule thus counts as paternalism in the sense that people limit their ability to act according to their will in a specific future situation. Although people consent to the intervention, the contractualist framework still amounts to paternalism since it allows for choice architectures that nudge individuals in directions they would probably not have otherwise chosen. In other words, although the contractualist framework provides a normative justification based on consent, individuals may still be nudged in a paternalistic manner.

In this section, I defended my contractualist approach for the justification of behavioural public policy from three potential objections. I first argued that the contractualist framework does not overcharge individuals with too much responsibility, since the state and the government must provide institutional circumstances that allow individuals to engage in collective choice rules and empower them to participate in bottom-up processes of policy-making. Second, I dismissed the claim that the contractualist approach must, like the welfarist approach it opposes, assume true and latent preferences. I argued that deliberative practices may lead to epistemic improvements and people may thereby reduce the effects of cognitive biases and protect themselves from (some) behavioural impairments. This by no means implies an inner rational agent or true preferences. And, lastly, I repudiated the objection that the contractualist approach would no longer count as paternalism. Instead, by limiting (future) freedom of choice, the contractualist framework still amounts to paternalism: paternalism towards oneself.

## Conclusion

In this paper, I aimed to clarify the fundamental normative implications of behavioural economics for the concept of libertarian paternalism, which is built on assumptions of conventional welfare economics. To this end, I demonstrated how substantial concerns arise for a normative approach of preference purification and informed or true preferences in the context of behavioural economics. Instead, I argued that we should pay greater attention to the fundamental normative dimension of behavioural public policy. I therefore introduced and defended a contractualist framework based on collective choice rules requiring individuals' participation and consent.

This paper emphasises crucial normative shortcomings of libertarian paternalism as a justification for behavioural interventions and proposes a new approach based people's engagement in collective choice rules. It may therefore contribute to untangling behavioural public policy from its normative justification based on libertarian paternalism. Ever since Sunstein and Thaler's original introduction, fundamental normative questions have been inextricably linked to their framework of libertarian paternalism and thus to their critical welfarist underpinning. However, since their normative reasoning is substantially flawed, it seems urgent that we clarify adequate normative approaches, given the increasing application of behavioural public policy across the world. To this end, a contractualist approach may offer a new normative perspective and help to further clarify the fundamental normative implications of behavioural economics.

Whether or not behavioural economics should be considered a return to the origins of economics as a discipline as invented by Adam Smith (Thaler 2016), it certainly advances more evidence-based methods and approaches. Yet, while behavioural economics was once considered a revolutionary and paradigm-shifting undertaking, today its aspirations sometimes seem rather moderate (Thaler 2016). However, I hope to have shown that behavioural economics has fundamental normative implications. Thus, while revealing the normative dimension of economics may lead back to Adam Smith, rediscovering its application for conventional, neoclassical economics today may help us to realise behavioural economics' radical and transformative potential (Spiegler 2017).

## Compliance with Ethical Standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References


Anand, Paul, and Alastair Gray. 2009. Obesity as market failure: Could a "deliberative economy" overcome the problems of paternalism? *Kyklos* 62 (2): 182–190. https://doi.org/10.1111/j.1467-6435.2009.00430.x.

Angner, Erik. 2015. To navigate safely in the vast sea of empirical facts: Ontology and methodology in behavioural economics. *Synthese* 192 (11). Springer Netherlands: 3557–3575. https://doi.org/10.1007/s11229-014-0552-9.

Arrow, Kenneth Joseph. 1963. *Social choice and individual values*. New Haven & London: Yale University Press.

Ashford, Elizabeth, and Tim Mulgan. 2012. Contractualism. In *The Stanford encyclopedia of philosophy (fall 2012 edition)*, ed. Edward N. Zalta.

Baldwin, Robert. 2014. From regulation to behaviour change: Giving nudge the third degree. *The Modern Law Review* 77 (6): 831–857. https://doi.org/10.1111/1468-2230.12094.

Banuri, Sheheryar, Stefan Dercon, and Varun Gauri. 2017. Biased policy professionals. *World Bank Policy Research Working Paper* 8113.

Barton, Adrien, and Till Grüne-Yanoff. 2015. From libertarian paternalism to nudging—And beyond. *Review of Philosophy and Psychology* 6 (3): 341–359. https://doi.org/10.1007/s13164-015-0268-x.

Behavioral Insights Team. 2017. *The Behavioural insights Team: Update report 2016–17*. London.

Berg, Nathan. 2003. Normative behavioral economics. *The Journal of Socio-Economics* 32 (4): 411–427. https://doi.org/10.1016/S1053-5357(03)00049-0.

Berg, Nathan, and Gerd Gigerenzer. 2010. As-if behavioral economics: Neoclassical economics in disguise. *History of Economic Ideas* XVIII (1): 133–165.

Bernheim, B.Douglas. 2009. Behavioral welfare economics. *Journal of the European Economic Association* 7 (2–3): 267–319. https://doi.org/10.1162/JEEA.2009.7.2-3.267.

Bernheim, B. Douglas. 2016. The good, the bad, and the ugly: A unified approach to behavioral welfare. *Journal of Benefit-Cost Analysis* 7 (1): 12–68. https://doi.org/10.1017/bca.2016.5.

Bernheim, B. Douglas, and Antonio Rangel. 2007. Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review* 97 (2): 464–470. https://doi.org/10.1257/aer.97.2.464.

Bernheim, B. Douglas, and Antonio Rangel. 2009. Beyond revealed preference: Choice theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics* 124 (1): 51–104. https://doi.org/10.1162/qjec.2009.124.1.51.

Binder, Martin. 2014. A constitutional paradigm is not enough —Would sovereign citizens really agree to manipulative nudges?—A reply to Christian Schubert. *Journal of Evolutionary Economics* 24 (5): 1115–1120. https://doi.org/10.1007/s00191-014-0377-1.

Bleichrodt, Han, Jose Luis Pinto, and Peter P. Wakker. 2001. Making descriptive use of Prospect theory to improve the prescriptive use of expected utility. *Management Science* 47 (11): 1498–1514. https://doi.org/10.1287/mnsc.47.11.1498.10248.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. Salience and consumer choice. *Journal of Political Economy* 121 (5): 803–843. https://doi.org/10.1086/673885.

Bovens, Luc. 2009. The ethics of nudge. In *Preference change: Approaches from philosophy, economics and psychology*, ed. Sven Ove Hansson and Till Grüne-Yanoff, 207–219. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-2593-7_10.

Buchanan, James M. 1975. *The limits of liberty. Between anarchy and leviathan*. Chicago: Chicago University Press.

Busemeyer, Jerome R., and James T. Townsend. 1993. Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review* 100 (3): 432–459.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. American economic association salience and taxation: Theory and evidence. *The American Economic Review* 99 (4): 1145–1177. https://doi.org/10.1257/aer.99.4.1145.

Cowen, Tyler. 1993. The scope and limits of preference sovereignty. *Economics and Philosophy* 9 (2). J. F. Kennedy Institute, Library: 253–269. https://doi.org/10.1017/S0266267100001553.

DellaVigna, Stefano. 2009. Psychology and economics: Evidence from the field. *Journal of Economic Literature* 47 (2): 315–372. https://doi.org/10.1257/jel.47.2.315.

Dold, Malte F. 2017. Back to Buchanan? Explorations of welfare and subjectivism in behavioral economics. Discussion Paper Series, *Wilfried-Guth-Stiftungsprofessur Für Ordnungs- Und Wettbewerbspolitik* 3.

Dold, Malte F., and Christian Schubert. 2016. Normative behavioral economics revisited: Toward a behavioral normative economics. *Working Paper*: 1–18.

Dworkin, Gerald. 1972. Paternalism. *The Monist* 56 (1): 64–84.

Dworkin, Gerald. 2017. Paternalism. In *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*, edited by Edward N. Zalta.

The Economist. 2008. *An Idea for Lent: Carrot and stickK*.

Elster, Jon. 1979. *Ulysses and the sirens : Studies in rationality and irrationality*. Cambridge: Cambridge University Press.

Elster, Jon. 1983. *Sour grapes: Studies in the subversion of rationality*. Cambridge: Cambridge University Press.

Ferey, Samuel. 2011. Paternalisme Libéral et Pluralité Du Moi. *Revue Économique* 62 (4): 737–750.

Fox, Craig R., and David Tannenbaum. 2015. The curious politics of the "nudge". *The New York Times, no.* (September 26).

Fumagalli, Roberto. 2016. Decision sciences and the new case for paternalism: Three welfare-related justificatory challenges. *Social Choice and Welfare* 47 (2). Springer Berlin Heidelberg: 459–480. https://doi.org/10.1007/s00355-016-0972-1.

Gauthier, David. 1986. *Morals by agreement*. Oxford: Oxford University Press.

Gigerenzer, Gerd. 2015. On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology* 6 (3): 361–383. https://doi.org/10.1007/s13164-015-0248-1.

Glaeser, Edward L. 2006. Paternalism and psychology. *University of Chicago Law Review* 73 (1): 133–156.

Goldin, Jacob. 2015. Which way to nudge? Uncovering preferences in the behavioral age. *Yale Law Journal* 125 (1): 226–270. https://doi.org/10.2139/ssrn.2570930.

Griffin, James. 1986. *Well-being: Its meaning, measurement and moral importance*. Oxford: Clarendon Press.

Grill, Kalle, and Danny Scoccia. 2015. Introduction. *Social Theory and Practice* 41 (4): 577–578. https://doi.org/10.5840/soctheorpract201541431.

Grüne-Yanoff, Till. 2009. *Welfare notions for soft paternalism. Papers on economics and evolution*. Vol. 917. Jena.

Grüne-Yanoff, Till. 2012. Old wine in new casks: Libertarian paternalism still violates Liberal principles. *Social Choice and Welfare* 38 (4): 635–645. https://doi.org/10.1007/s00355-011-0636-0.

Guala, Francesco, and Luigi Mittone. 2015. A political justification of nudging. *Review of Philosophy and Psychology* 6 (March): 385–395. https://doi.org/10.1007/s13164-015-0241-8.

Hacker, Philipp. 2016a. Nudge 2.0: The future of behavioural analysis of law in Europe and beyond. *European Review of Private Law* 24 (2): 297–322.

Hacker, Philipp. 2016b. Nudging and autonomy. A philosophical and legal appraisal. In *Handbook of research methods in consumer law*, ed. Hans-W. Micklitz, Kai Purnhagen, and Anne-Lise Sibony. Cheltenham, United Kingdom: Edward Elgar Publishing.

Hagman, William, David Andersson, Daniel Västfjäll, and Gustav Tinghög. 2015. Public views on policies involving nudges. *Review of Philosophy and Psychology* 6 (3): 439–453. https://doi.org/10.1007/s13164-015-0263-2.

Halpern, David. 2016. *Inside the nudge unit: How small changes can make a big difference*. London: Random House.

Hansen, Pelle Guldborg, and Andreas Maaløe Jespersen. 2013. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation* 4 (1): 3–28. https://doi.org/10.2307/2489305.

Harsanyi, John C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63 (4): 309–321.

Harsanyi, John C. 1977. Rule utilitarianism and decision theory. *Erkenntnis* 11 (1): 25–53. https://doi.org/10.1007/BF00169843.
Hausman, Daniel M. 2012. *Preference, value, choice, and welfare*. Cambridge, United Kingdom: Cambridge University Press.
Hausman, Daniel M. 2016. On the econ within. *Journal of Economic Methodology* 23 (1). Routledge: 26–32. https://doi.org/10.1080/1350178X.2015.1070525.
Hausman, Daniel M., and Brynn Welch. 2010. Debate: To nudge or not to nudge. *Journal of Political Philosophy* 18 (1): 123–136. https://doi.org/10.1111/j.1467-9760.2009.00351.x.
Hausman, Daniel M., Michael S. McPherson, and Debra Satz. 2017. *Economic analysis, moral philosophy, and public policy*. New York: Cambridge University Press.
Heath, Joseph. 2014. *Morality, competition, and the firm*. Oxford and New York: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199990481.001.0001.
Hédoin, Cyril. 2016. Normative economics and paternalism: The problem with the preference-satisfaction account of welfare. *Constitutional Political Economy, October*: 1–25. https://doi.org/10.1007/s10602-016-9227-5.
Heidbrink, Ludger. 2015. Libertarian paternalism, sustainable self-binding and bounded freedom. In *The politics of sustainability. Philosophical perspectives*, edited by Dieter Birnbacher and May Torseth, 173–194. Abingdon & New York: Routledge. https://doi.org/10.4324/9781315721200.
Heidbrink, Ludger, and Johannes Reidel. 2011. Nachhaltiger Konsum Durch Politische Selbstbindung. *GAIA* 20 (3): 152–156.
Heidl, Stefan. 2016. *Philosophical problems of behavioral economics*. London & New York: Routledge.
Herzog, Lisa. 2008. Economic ethics for real humans - the contribution of behavioral economics to economic ethics. *Zeitschrift Für Wirtschafts- Und Unternehmensethik* 9: 112–128.
Hirshleifer, David, and Siew Hong Teoh. 2017. How psychological Bias shapes accounting and financial regulation. *Behavioural Public Policy* 1 (1): 87–105. https://doi.org/10.1017/bpp.2016.5.
Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016a. On the econ within: A reply to Daniel Hausman. *Journal of Economic Methodology* 23 (1). Routledge: 33–37. https://doi.org/10.1080/1350178X.2015.1070526.
Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016b. Preference purification and the inner rational agent: A critique of the conventional wisdom of Behavioural welfare economics. *Journal of Economic Methodology* 23 (1). Routledge: 1–25. https://doi.org/10.1080/1350178X.2015.1070527.
John, Peter. 2013. Experimentation, behaviour change and public policy. *The Political Quarterly* 84 (2): 238–246.
John, Peter, Sarah Cotterill, Hanhua Liu, Liz Richardson, Alice Moseley, Hisako Nomura, Graham Smith, Gerry Stoker, and Corinne Wales. 2011. Nudge, nudge, think, *think*. London & New York: Bloomsbury Publishing PLC. https://doi.org/10.5040/9781849662284.
Jung, Janice Y., and Barbara A. Mellers. 2016. American attitudes toward nudges. *Judgment and Decision Making* 11 (1): 62–74.
Just, David R. 2017. The behavioral welfare paradox: Practical, ethical and welfare implications of nudging. *Agricultural and Resource Economics Review* 46 (1): 1–20. https://doi.org/10.1017/age.2017.2.
Kahneman, Daniel. 1996. Comment (on Plott). In *The rational foundations of economic behaviour*, ed. K. Arrow, E. Colombatto, M. Perlmann, and C. Schmidt, 251–254. Basingstoke.
Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2010. Self-control and the development of work arrangements. *American Economic Review* 100 (2): 624–628. https://doi.org/10.1257/aer.100.2.624.
Kőszegi, Botond, and Matthew Rabin. 2007. Mistakes in choice-based welfare analysis. *The American Economic Review* 97 (2): 477–481.
Kőszegi, Botond, and Matthew Rabin. 2008. Choices, situations, and happiness. *Journal of Public Economics* 92 (8–9): 1821–1832. https://doi.org/10.1016/j.jpubeco.2008.03.010.
Lecouteux, Guilhem. 2015a. *Reconciling normative and behavioural economics. Thèse, École Doctorale de L'École Polytechnique*. Paris.
Lecouteux, Guilhem. 2015b. In search of lost nudges. *Review of Philosophy and Psychology* 6 (3): 397–408. https://doi.org/10.1007/s13164-015-0265-0.
Lecouteux, Guilhem. 2017. Stefan Heidl: Philosophical problems of Behavioural economics. *Œconomia* 7 (1): 143–148.
Lepenies, Robert, and Magdalena Małecka. 2015. The institutional consequences of nudging – Nudges, politics, and the law. *Review of Philosophy and Psychology* 6 (3): 427–437. https://doi.org/10.1007/s13164-015-0243-6.
Lepenies, Robert, and Magdalena Małecka. 2016. Nudges, Recht Und Politik: Institutionelle Implikationen. *Zeitschrift Für Praktische Philosophie* 3 (1): 487–530.

Luetge, Christoph. 2005. Economic ethics, business ethics and the idea of mutual advantages. *Business Ethics: A European Review* 14 (2): 108–118. https://doi.org/10.1111/j.1467-8608.2005.00395.x.

Luetge, Christoph, Thomas Armbrüster, and Julian Müller. 2015. Order ethics: Bridging the gap between contractarianism and business ethics. *Journal of Business Ethics*, July 22. https://doi.org/10.1007/s10551-015-2977-6.

Lynch, John G., and Gal Zauberman. 2006. When do you want it? Time, decisions, and public policy. *Journal of Public Policy & Marketing* 25 (1): 67–78. https://doi.org/10.1509/jppm.25.1.67.

McPherson, Michael S., and Matthew A. Smith. 2008. Nudging for liberty: Values in libertarian paternalism. *SSRN Electronic Journal*, no. January. https://doi.org/10.2139/ssrn.1220323.

McQuillin, Ben, and Robert Sugden. 2012a. Reconciling normative and behavioural economics: The problems to be solved. *Social Choice and Welfare* 38 (4): 553–567. https://doi.org/10.1007/s00355-011-0627-1.

McQuillin, Ben, and Robert Sugden. 2012b. How the market responds to dynamically inconsistent preferences. *Social Choice and Welfare* 38 (4): 617–634. https://doi.org/10.1007/s00355-011-0628-0.

Moles, Andrés. 2015. Nudging for Liberals. *Social Theory and Practice* 41 (4): 644–667. https://doi.org/10.5840/soctheorpract201541435.

Moseley, Alice, and Gerry Stoker. 2013. Nudging citizens? Prospects and pitfalls confronting a new heuristic. *Resources, Conservation and Recycling* 79 (October): 4–10. https://doi.org/10.1016/j.resconrec.2013.04.008.

Nys, Thomas R.V., and Bart Engelen. 2017. Judging nudging: Answering the manipulation objection. *Political Studies* 65 (I): 199–214. https://doi.org/10.1177/0032321716629487.

OECD. 2010. *Consumer policy toolkit*. OECD Publishing. https://doi.org/10.1787/9789264079663-en.

OECD. 2014. *Recommendation on consumer policy decision making*. Paris: OECD Publishing. https://doi.org/10.1787/9789264079663-en.

OECD. 2017a. *Use of Behavioural insights in consumer policy*. Paris. https://doi.org/10.1787/c2203c35-en.

OECD. 2017b. *Behavioural insights and public policy: Lessons from around the world*. Paris: OECD Publishing. https://doi.org/10.1787/9789264270480-en.

Pasche, Markus. 2014. Soft paternalism and nudging – Critique of the behavioral foundations. *MPRA Paper* No. 61140 (2116): 1–9.

Pedersen, Sofie Kragh, Alexander K. Koch, and Julia Nafziger. 2014. Who wants paternalism? *Bulletin of Economic Research* 66 (S1): S147–S166. https://doi.org/10.1111/boer.12030.

Pinto-Prades, Jose-Luis, and Jose-Maria Abellan-Perpiñan. 2012. When normative and descriptive diverge: How to bridge the difference. *Social Choice and Welfare* 38 (4): 569–584. https://doi.org/10.1007/s00355-012-0655-5.

Qizilbash, Mozaffar. 2012. Informed desire and the ambitions of libertarian paternalism. *Social Choice and Welfare* 38 (4): 647–658. https://doi.org/10.1007/s00355-011-0620-8.

Railton, Peter. 1986. Facts and values. *Philosophical Topics* 14 (2): 5–30.

Rawls, John. 1971. *A theory of justice*. Cambridge, Massachusetts: Harvard University Press.

Rebonato, Riccardo. 2012. *Taking liberties: A critical examination of libertarian paternalism*. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230391567.

Reisch, Lucia A., and Cass R. Sunstein. 2016. Do Europeans Like Nudges? *Judgment and Decision Making* 11 (4): 310–325 doi:https://doi.org/10.2139/ssrn.2739118.

Reisch, Lucia A., Cass R. Sunstein, and Wencke Gwozdz. 2017. Beyond carrots and sticks: Europeans support health nudges. *Food Policy* 69 (May): 1–10. https://doi.org/10.1016/j.foodpol.2017.01.007.

Reiss, Julian. 2013. *Philosophy of economics: A contemporary introduction*. New York and London: Routledge.

Rizzo, Mario J. 2017. Rationality - what? Misconceptions of neoclassical and behavioral economics. *Working Paper*: 1–24. https://doi.org/10.2139/ssrn.2927443.

Rizzo, Mario J., and Douglas Glen Whitman. 2009. The knowledge problem of the new paternalism. *Brigham Young University Law Review* 2009 (4): 905–968.

Rubinstein, Ariel, and Yuval Salant. 2012. Eliciting welfare preferences from Behavioural data sets. *Review of Economic Studies* 79 (2012): 375–387. https://doi.org/10.1093/restud/rdr024.

Salant, Yuval, and Ariel Rubinstein. 2008. (A , F): Choice with frames. *Review of Economic Studies* 75 (4): 1287–1296. https://doi.org/10.1111/j.1467-937X.2008.00510.x.

Scanlon, Thomas M. 1998. *What we owe to each other*. Cambridge, Massachusetts: Harvard University Press.

Schaal, Gary S., and Claudia Ritzi. 2009. Rationale Selbstbindung Und Die Qualität Politischer Entscheidungen – Liberale Und Deliberative Perspektiven. In *Techniken Rationaler Selbstbindung*, ed. Gary S. Schaal, 55–74. Berlin: LIT Verlag.

Schaal, Gary S. 2009. Techniken rationaler Selbstbindung. Berlin: LIT Verlag.

Schnellenbach, Jan. 2011. Wohlwollendes Anschubsen: Was Ist Mit Liberalem Paternalismus Zu Erreichen Und Was Sind Seine Nebenwirkungen? *Perspektiven der Wirtschaftspolitik* 12 (4): 445–459. https://doi.org/10.1111/j.1468-2516.2012.00381.x.

Schnellenbach, Jan. 2012. Nudges and norms: On the political economy of soft paternalism. *European Journal of Political Economy* 28 (2). Elsevier B.V.: 266–277. doi:https://doi.org/10.1016/j.ejpoleco.2011.12.001.
Schnellenbach, Jan, and Christian Schubert. 2015. Behavioral political economy: A survey. *European Journal of Political Economy* 40 (December): 395–417. https://doi.org/10.1016/j.ejpoleco.2015.05.002.
Schramme, Thomas. 2016. Die Politische Quacksalberei Des Libertären Paternalismus. *Zeitschrift Für Praktische Philosophie* 3 (1): 531–558.
Schubert, Christian. 2012a. Pursuing Happiness. *Kyklos* 65 (2): 245–261. https://doi.org/10.1111/j.1467-6435.2012.00537.x.
Schubert, Christian. 2012b. Is novelty always a good thing? Towards an evolutionary welfare economics. *Journal of Evolutionary Economics* 22 (3): 585–619. https://doi.org/10.1007/s00191-011-0257-x.
Schubert, Christian. 2014. Evolutionary economics and the case for a constitutional libertarian paternalism—A comment on Martin Binder, "should evolutionary economists embrace libertarian paternalism?". *Journal of Evolutionary Economics* 24 (5): 1107–1113. https://doi.org/10.1007/s00191-014-0379-z.
Schubert, Christian. 2015a. What do we mean when we say that innovation and entrepreneurship (policy) increase "welfare"? *Journal of Economic Issues* 49 (1): 1–22. https://doi.org/10.1080/00213624.2015.1013859.
Schubert, Christian. 2015b. Opportunity and preference learning. *Economics and Philosophy* 31 (2): 275–295. https://doi.org/10.1017/S0266267115000139.
Schubert, Christian. 2017. Exploring the (Behavioural) political economy of nudging. *Journal of Institutional Economics* 13 (3): 499–522. https://doi.org/10.1017/S1744137416000448.
Schubert, Christian, and Christian Cordes. 2013. Role models that make you unhappy: Light paternalism, social learning, and welfare. *Journal of Institutional Economics* 9 (2): 131–159. https://doi.org/10.1017/S1744137413000015.
Sen, Amartya K. 1970. *Collective choice and social welfare*. San Francisco: Holden-Bay.
Sen, Amartya. 1987. *On ethics and economics*. Oxford and New York: Basil Blackwell.
Sibony, Anne-Lise, and Alberto Alemanno. 2015. The emergence of behavioural policy-making: A European perspective. In *Nudge and the law: A European perspective*, edited by Alberto Alemanno and Anne-Lise Sibony, 1–25. Oxford and Portland, Oregon: Hart Publishing.
Smith, Matthew A., and Michael S. McPherson. 2009. Nudging for equality: Values in libertarian paternalism. *Administrative Law Review* 61 (2): 323–342.
Sobel, David. 1994. Full information accounts of well-being. *Ethics* 104: 784–810.
Social and Behavioral Sciences Team. 2016. *Social and behavioral sciences Team:* 2016 *Annual Report*. Washington, D.C.: Executive Office of the President, National Science and Technology Council.
Spiegler, Ran. 2017. Behavioral economics and the atheoretical style. *CEPR Discussion Paper* DP11786: 1–28.
Sugden, Robert. 2004. The opportunity criterion: Consumer sovereignty without the assumption of coherent preferences. *American Economic Review* 94 (4): 1014–1033. https://doi.org/10.1257/0002828042002714.
Sugden, Robert. 2006. Taking unconsidered preferences seriously. *Royal Institute of Philosophy Supplement* 59 (1): 209–232. https://doi.org/10.1017/S1358246106059108.
Sugden, Robert. 2007. The value of opportunities over time when preferences are unstable. *Social Choice and Welfare* 29 (4): 665–682. https://doi.org/10.1007/s00355-007-0250-3.
Sugden, Robert. 2008. Why incoherent preferences do not justify paternalism. *Constitutional Political Economy* 19 (3): 226–248. https://doi.org/10.1007/s10602-008-9043-7.
Sugden, Robert. 2009. On nudging: A review of nudge: Improving decisions about health, wealth and happiness by Richard H. Thaler and Cass R. Sunstein. *International Journal of the Economics of Business* 16 (3): 365–373. https://doi.org/10.1080/13571510903227064.
Sugden, Robert. 2010. Opportunity as mutual advantage. *Economics and Philosophy* 26 (1): 47. https://doi.org/10.1017/S0266267110000052.
Sugden, Robert. 2013. The behavioural economist and the social planner: to whom should behavioural welfare economics be addressed? *Inquiry* 56 (5): 519–538. https://doi.org/10.1080/0020174X.2013.806139.
Sugden, Robert. 2015a. Looking for a psychology for the inner rational agent. *Social Theory and Practice* 41 (4): 579–598. https://doi.org/10.5840/soctheorpract201541432.
Sugden, Robert. 2015b. Opportunity and preference learning: A reply to Christian Schubert. *Economics and Philosophy* 31 (2): 297–303. https://doi.org/10.1017/S0266267115000140.
Sugden, Robert. 2016. *Do people really want to be nudged towards healthy lifestyles? International Review of Economics, September*. Springer Berlin Heidelberg. https://doi.org/10.1007/s12232-016-0264-1.
Sugden, Robert. 2017. Better off, as judged by themselves: A reply to cass sunstein. *International Review of Economics*, July. Springer Berlin Heidelberg, 1–5. doi:https://doi.org/10.1007/s12232-017-0281-8.
Sugden, Robert. 2018. *The community of advantage: A behavioural economist's defence of the market*. Oxford & New York: Oxford University Press.

Sunstein, Cass R. 2016. People prefer system 2 nudges (kind of). *Duke Law Journal* 66: 121–168. https://doi.org/10.2139/ssrn.2731868.
Sunstein, Cass R. 2017a. Do people like nudges? *Administrative Law Review* 68 (2): 177.
Sunstein, Cass R. 2017b. *Human agency and behavioral economics: Nudgung fast and slow*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-55807-3.
Sunstein, Cass R. 2017c. Better off, as judged by themselves: A comment on evaluating nudges. *International Review of Economics*, June. Springer Berlin Heidelberg, Forthcoming. doi:https://doi.org/10.1007/s12232-017-0280-9.
Sunstein, Cass R., and Richard H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review* 70 (4): 1159–1202. https://doi.org/10.2307/1600573.
Sunstein, Cass R., and Richard H. Thaler. 2006. Preferences, paternalism, and liberty. In *Preferences and well-being. Royal Institute of philosophy supplement 59*, ed. Serena Olsaretti, 233–264. Cambridge: Cambridge University Press. https://doi.org/10.1017/S135824610605911X.
Sunstein, Cass R., Lucia A. Reisch, and Julius Rauber. 2017. 'Behavioral Insights All Over the World? Public Attitudes Toward Nudging in a Multi-Country Study', 1–31.
Tannenbaum, David, Craig R. Fox, and Todd Rogers. 2017. On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour* 1 (7). Macmillan Publishers Limited, part of Springer Nature.: 130. https://doi.org/10.1038/s41562-017-0130.
Thaler, Richard H. 2016. Behavioral economics: Past, present, and future. *Article Based on the Presidential Address Given at the American Economic Association Annual Meeting in January* 2016.
Thaler, Richard H., and Shlomo Benartzi. 2004. Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy* 112 (S1): S164–S187. https://doi.org/10.1086/380085.
Thaler, Richard H., and Cass R. Sunstein. 2003. Libertarian Paternalism. *The American Economic Review* 93 (2): 175–179.
Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven & London: Yale University Press.
van Aaken, Anne. 2006. Begrenzte Rationalität Und Paternalismusgefahr: Das Prinzip Des Schonendsten Paternalismus. In *Paternalismus Und Recht*, edited by Michael Anderheiden, Peter Bürkli, Hans-Michael Heinig, Stephan Kirste, and Kurt Seelmann, 109–144. Tübingen: Mohr Siebeck.
van Aaken, Anne. 2015. Das Deliberative Element Juristischer Verfahren Als Instrument Zur Überwindung Nachteiliger Verhaltensanomalien. In *Recht Und Verhalten: Beiträge Zu Behavioral Law and Economics*, edited by Christoph Engel, Markus Englerth, Jörn Lüdemann, and Indra Spiecker genannt Döhmann, 189–230. Tübingen: Mohr Siebeck.
Vringer, Kees, Herman R.J. Vollebergh, Daan van Soest, Eline van der Heijden, and Frank Dietz. 2015. *Sustainable consumption dilemmas. OECD environment working papers*. Paris. https://doi.org/10.1787/5js4k112t738-en.
White, Mark D. 2013. *The manipulation of choice*. New York: Palgrave Macmillan US. https://doi.org/10.1057/9781137313577.
Whitman, Douglas Glen, and Mario J. Rizzo. 2015. The problematic welfare standards of behavioral paternalism. *Review of Philosophy and Psychology* 6 (3): 409–425. https://doi.org/10.1007/s13164-015-0244-5.



**Johann Jakob Häußermann** works at the Center for Responsible Research and Innovation at the Fraunhofer-Institute for Industrial Engineering and is currently completing his PhD at the TUM School of Governance. He is particularly interested in the interface between innovation and ethics and develops new approaches to shape new technologies and business models responsibly and for the benefit of society. Most recently, he focused on the ethical implications of artificial intelligence, how such systems can be responsibly developed and used to add value to society.