

Would you choose it again? On confirmed choices as a proxy of welfare*

João V. Ferreira[†]

This version: 11/2018

Abstract

It has been standard practice in neoclassical economics to equate choice with welfare. Following the evidence on bounded rationality, problems of self-control, among other arational behaviours, several authors have proposed models that equate some (as opposed to all) choices with welfare. In this paper, I first distinguish two prominent criteria in behavioural economics that are based on choices that remain constant across context and time (*context-independent choices*) and choices that are made after rational deliberation (*reason-based choices*). I then propose an alternative criterion based on choices that individuals confirm after reflecting upon their behaviour (*confirmed choices*). I articulate three complementary arguments that, I claim, support confirmed choices as a reliable (though fallible) proxy of welfare. Finally, I discuss the implications of this view and contrast it to fully-informed hypothetical accounts of well-being and libertarian paternalism.

JEL classification: D01; D60; D90; H40.

Keywords: Choice; Welfare; Confirmed choices; Behavioural welfare economics; Libertarian paternalism.

1. Introduction

It has been standard practice in neoclassical economics to assume a tight link between choice and individual welfare. This link has been justified on different grounds: sometimes choices are said to reveal preferences that individuals want satisfied, other times choices are said to be a good proxy of subjective well-being, and still other that individuals want choices satisfied even

*I am grateful to Jose Apesteguia, Salvador Barberà, Koen Decancq, Marc Fleurbaey, Nicolas Gravel, Fabrice Le Lec, Stephane Lemaire, Olivier L'Harridon, Stephane Luchini, Serena Olsaretti, Erik Schokkaert, Arthur Schram, Robert Sugden, Benoît Tarrow, Miriam Teschl, Alain Trannoy, Andrew Williams, and the participants of the ASSET 2017, the Workshop in Individual Choice and Freedom, and the LAGV 2018 for valuable discussions, comments and suggestions.

[†]Univ Lyon, Université Lumière Lyon 2, GATE L-SE UMR 5824, F-69130 Ecully, France. Email: ferreira@gate.cnrs.fr

if they do not reveal some underlying objectives.¹ According to any of these views it is possible to derive from individual choices a ranking of alternatives, from “better” to “worse”, that can be used to evaluate the desirability of different states of affairs.

The link between choice and welfare has traditionally relied on the assumption that individual actual choices are consistent over time and across contexts. Evidence from psychology and behavioural economics on bounded rationality, framing effects, problems of self-control, among other *arational* behaviours show however that choice behaviour is often at odds with the latter assumption.² For example, there is by now considerable evidence that people’s behaviour can be influenced by environmental cues and “anchors” (see e.g. Tversky and Kahneman 1974; Ariely et al. 2003; Choi et al. 2004): Even major decisions about retirement savings can be determined by whether some options, such as membership of a saving plan, are described as a default option from which one can opt out or as an option to which one opts in. These kinds of findings question if an individual’s actual choices provide a reliable welfare ranking of alternatives. To illustrate this problem, consider the following example:

The snacks choice. Suppose that Norah is offered the choice between an *apple* and a *snickers* bar. In a first situation she is asked to choose in advance which one she wants to consume one week later, and she chooses the *apple*. One week later she is asked which one she wants to consume immediately, and she now chooses the *snickers*.³

Which of these choices, if either, provide a good indication of what is in Norah’s best interest? In this paper I combine insights from behavioural economics and moral philosophy to assess the merits of different kinds of choices as proxies of welfare.⁴ First, I review two prominent criteria in behavioural economics that equate some (as opposed to all) choices with welfare. The first

¹See e.g. Little (1949), Samuelson (1963), and Gul and Pesendorfer (2008) for a mix of these claims (see Bernheim 2009, pp. 290-3 for a review). For example, Little (1949, p. 98) argues that one could say that “a person is, on the whole, likely to be happier the more he can have what he would choose. Or, alternatively, one can say that it is a good thing that he should be able to have what he would choose”.

²The literature is vast. See e.g. Kahneman (2011), Rabin (2013), and Hoff and Stiglitz (2016) for reviews. I borrow the term “arational” from Sugden (2006, p. 216) to note that many of the behaviours documented in the literature are not necessarily *irrational*.

³This example is taken from the experiment by Read and van Leeuwen (1998). In the experiment subjects are not aware that they can redo their choice a week later, and between 62% to 82% of subjects (depending on the treatment that differ in terms of current and future states of hunger) that chose a “healthy” snack in the advance choice reversed their choice for an “unhealthy” snack in the immediate choice.

⁴I do not directly address the question *if* choices are the “right” proxy of welfare (cf. Broome 1978). Many authors believe otherwise, and some have proposed alternative proxies of welfare based on individuals’ experiences of pleasure and pain or the activation patterns of specific areas in the brain (see e.g. Kahneman et al. 1997 and Camerer et al. 2004 respectively; see Fumagalli 2013 for a critical review). Using choice behaviour (or stated choices) to make welfare comparisons between different alternatives is, however, a common practice in economic welfare analysis that is probably to last. In fact, the balance of evidence seems to point towards complementarity rather than substitutability with other measures. There is then a pragmatic reason to identify which choices are a reliable proxy of welfare (see also Chambers and Hayashi 2012 and Manzini and Mariotti 2014). Note that this aim is related but independent from the aim of identifying which preferences have (or ought to have) prudential value (see e.g. Olsaretti 2006). For instance, it is coherent to uphold that some hypothetical preferences ought to have prudential value, but that in general actual choices (or preferences) are the best proxy of welfare available (see e.g. Arneson 1990, p. 164).

of these contends that only choices that remain constant across context and time (*context-independent choices*), irrespective of their origin, are good proxies of welfare. In the *snacks choice* example, this criterion is agnostic about which snack is “better” for Norah. The second contends that only choices that are made after rational deliberation (*reason-based choices*) are reliable proxies of welfare. This approach demands information about the origin of choices in order to evaluate which ones are reliable proxies of welfare. In the *snacks choice* example, suppose we have gathered evidence that immediate consumption triggers impulsiveness. Most versions of this criterion would then deem the apple better for Norah.

I argue that both these approaches miss an important distinction between choices that individuals would confirm and choices that individuals would do not confirm. An individual is said to confirm her own choice if, when reflecting upon facing that choice again, she would not want herself to choose differently at the time of reflection. I call such “self-authenticated” choice a *confirmed choice*. To illustrate this view, consider the following example:

The smoking choice. Suppose that Norah is offered the choice between *smoking* and *not smoking*. In a first situation Norah is asked to choose after a period of abstinence whether or not to smoke and she chooses *smoking*. In a second situation she is asked to choose immediately after having smoked one cigarette and she chooses again *smoking*. Moreover, suppose that Norah has stated that if faced with the same choices again she would want herself to choose *not smoking*.⁵

I argue that the last piece of information is essential to make reliable welfare comparisons from choice behaviour. It reveals that Norah has not behaved according to how, upon forming a reflected judgement about her behaviour, she would want herself to behave. Using confirmations as a refinement of a general framework for describing choice behaviour recently proposed by Bernheim and Rangel (2009), I show that a criterion based on confirmed choices is agnostic about which of the two alternatives, *smoking* or *not smoking*, is better for Norah. I will also show that in other cases this criterion can discern which of two alternatives should be seen as welfare superior when the other two criteria cannot.

I provide three complementary arguments that support confirmed choices as a reliable (though fallible) proxy of welfare. First, I argue that respecting confirmed choices is conducive to the satisfaction of individual preferences. I distinguish preferences that individuals want satisfied from those they do not, and argue that empirical data on confirmations is necessary to identify which choices reveal welfare-enhancing preferences. Second, I argue that confirmed choices are a more accurate proxy of subjective well-being than context-independent and reason-based

⁵Smoking is a typical example of a conflict between people’s behaviour and their judgements about their behaviour that involves high stakes (see e.g. World Health Organization 2018). According to a recent report from the U.K.’s Office for National Statistics (2017), in Great Britain 60.8% of people aged 16 years and above who currently smoke said they wanted to quit. Similar figures can be found in other countries (see e.g. Center for Disease Control and Prevention 2011 for the U.S.).

choices. This, I claim, follows not only from the previous argument but also from the fact that confirmations trace important aspects of (multidimensional) subjective well-being such as living according to personal values and being *who* one wants to be. Third, I argue that respecting confirmed choices is conducive to the respect of individual sovereignty. The underlying reason is that individual sovereignty should, I claim, not only concern individuals' choices but also their judgements about their choices. These arguments, when taken together, provide an additional liberal reason to satisfy individual choices conditional on the fact that they are confirmed.

To avoid misunderstandings it is worth emphasizing that I do regard confirmed choices as a fallible proxy of welfare.⁶ People often lack relevant information, change their minds, are prone to cognitive bias, and/or adapt to their opportunities. The reliability of choices and subjective attitudes as proxies of welfare is conditional, among other things, on the information and competence of the individuals (Hausman and McPherson 2009; Hausman 2012). To avoid repetition I will assume that minimal conditions are met, and return to this issue in section 5. At that point, I contrast my view to fully-informed hypothetical accounts of well-being and argue that even if one endorses such a notion of well-being, confirmations are still necessary in order to make reliable welfare comparisons from choice behaviour. Finally, I discuss some of the implications of this view using policy examples and compare them to libertarian paternalistic interventions.

Before proceeding, it is worth noting that this approach is related to what philosophers often call second-order preferences, desires, or volitions.⁷ For example, "I want not to want to smoke" is a second-order desire. I distinguish from previous authors by focusing on asynchronous judgements about behaviour, and by relating to (behavioural) welfare economics. This approach is also related but different from the traditional view of individual meta-preferences adopted in economics (e.g. Sen 1977; George 1984). In general, meta-preferences are assumed to be a single ordering of multiple preferences defined over the universe of alternatives. Here choices, as opposed to multiple preferences, are the primitive for the individuals' judgements. Likewise, confirmed choices are related but different to the standard interpretation of meta-choices (see e.g. Bernheim 2009, p. 311). While a choice is said to be confirmed as long as the individual would not want herself to choose differently when reflecting upon facing the same choice again, a meta-choice is assumed to be an advance choice between two or more choices. Finally, the notion of confirmed choice resembles the notion of rational choice proposed by Gilboa and Schmeidler (2001, pp. 17-8): "An action, or a sequence of actions is rational for a decision maker if,

⁶I consider welfare to correspond to the agent's interest *as judged by herself*. By this I mean her actual judgement, as opposed to a judgement of her hypothetical fully-informed rational self (e.g. Sunstein and Thaler 2003). This means that a proxy of welfare is fallible only if it is recognized to be so by the agent at a given point in time. This accords with the tradition in economics to treat individual judgements as decisive in assessing the relative welfare associated with different alternatives.

⁷See Frankfurt (1971) for the philosophical basis of higher-order desires and volitions. See Jeffrey (1974) for an early treatment of second-order preferences. Confirmation, even if considerably less demanding, also relates to some concepts of identification (e.g. Frankfurt 1988, Noggle 1999; cf. Watson 1975, Bratman 1996, 2003, Scanlon 2002, and Lippert-Rasmussen 2003).

when the decision maker is confronted with an analysis of the decisions involved, but with no additional information, she does not regret her choices” (see also Gilboa 2010).⁸ A confirmed choice, however, is said to be confirmed based on a judgement that can be or not informed by an analysis of the decisions involved and/or additional information. Confirmation also differs from regret. It can be said to represent, instead, the answer to the following thought experiment ingrained in common wisdom: “From where you stand now, would you choose it again?”

The remaining of the paper is organized as follows. In section 2 I present the criterion based on context-independent choices and the formal framework that is used to illustrate the welfare inferences of the different criteria. In section 3 I present the criterion based on reason-based choices. In section 4, the core of the paper, I propose an alternative based on confirmed choices and articulate three complementary arguments in its favour. I then discuss the implications of this view and contrast it to fully-informed hypothetical accounts of well-being and libertarian paternalism (section 5). Section 6 concludes.

2. Context-independent Choices

There are several proposals of how to reconcile welfare economics with the behavioural findings that actual choices are often “inconsistent” and possibly opposed to agents’ interests. One of the closest in spirit to traditional welfare economics assumes that only those choices, among all possibly context-dependent choices, that remain constant across context and time (*context-independent choices*) are reliable proxies of welfare.

This view has been recently formalized by Bernheim and Rangel (2009) (hereinafter B&R).⁹ They consider a general framework for describing choice behaviour in which X denotes the set of all possible *alternatives* such as consumption bundles or any other state of affairs, as long as alternatives are complete and mutually exclusive descriptions of the world. In order to model context-dependent behaviour, B&R define a *generalized choice situation* (GCS), denoted $G = (A, d)$, as the combination of a standard *choice situation* $A \subseteq X$ (i.e., an opportunity set of alternatives) and an *ancillary condition* d . An ancillary condition can be the manner in which information is presented, the time period at which a decision is taken, or other frames as long as it is considered to be a “feature of the choice environment that may affect behaviour, but [that is] not taken as relevant to a social planner’s evaluation” (B&R, p. 55). According to this interpretation, the *snacks choice* example can be represented by treating snacks as alternatives

⁸Note that contrary to Gilboa and Schmeidler (2001) I am not concerned with which choices should be seen as rational. Similarly, I make no claim concerning whether agents have (or ought to have) normative reasons to behave in one or other way (cf. Scanlon 2002, Sobel 2009, and Parfit 2011, Part 1). As argued by Broome (1978, p. 324), when considering an “authority” that may act on behalf of an agent or choose policies that affect an agent it is sufficient to impute to the authority a desire to respect the agent’s interest.

⁹See also Bernheim and Rangel (2007) and Bernheim (2009). See Salant and Rubinstein (2008) for an analogous framework developed to represent the impact of “frames” on choice behaviour. See Manzini and Mariotti (2014) for a criticism of B&R based on the fact that their approach does not rely on an explicit model of decision making. See Chambers and Hayashi (2012) and Nishimura (2018) for other “model-less approaches” that rely upon choice behaviour to make welfare comparisons.

and the distance to consumption as an ancillary condition.

Let \mathcal{G}^* denote the set of all generalized choice situations contemplated by an (active) *observer*.¹⁰ Individual behaviour is modelled through a choice correspondence $C : \mathcal{G}^* \Rightarrow X$, that assigns a set of alternatives $C(A, d) \subseteq A$ to every generalized choice situation $(A, d) \in \mathcal{G}^*$.¹¹ An alternative $x \in C(A, d)$ is interpreted as an option that the agent *selects* and is willing to choose when facing (A, d) . Welfare analysis can then be performed by defining a welfare binary relation, P , where xPy means that x is “better than” y . B&R’s preferred welfare ranking of alternatives is based on what they call an “*unambiguous choice* relation”, denoted P^* , and defined as follows:

$$xP^*y \text{ if and only if for all } (A, d) \in \mathcal{G}^* \text{ such that } x, y \in A, \text{ we have } y \notin C(A, d). \quad (1)$$

In other words, x is said to be better than y if and only if y is never selected when x is available. Since by assumption every subset of X , including $\{x, y\}$, is in the domain \mathcal{G}^* (see fn. 11), this means that x is only said to be better than y when x is chosen at least once over y and y is never selected when x is available. Thus, only choices between two alternatives that remain stable for all observed choice situations and ancillary conditions determine the welfare ranking of different alternatives. The welfare ranking P^* is acyclic under these assumptions, which guarantees the existence of welfare optima for any choice situation and the identification of welfare improvements.

It is worth noting that B&R are agnostic about the process that gives rise to choices. Contrary to other authors (e.g. Rubinstein and Salant 2012), they do not assume the existence of an underlying context-independent stable preference that can be reconstructed by eliminating mistakes. The presumption is that “choices provide appropriate guidance because they are choices” (B&R, p. 52; see also Bernheim 2009, pp. 290-3). Their criterion seems like a natural extension of a choice satisfaction principle to settings in which context-dependent behaviour is prevalent.

On the one hand, this approach is appealing since it relies exclusively on choice data. On the other hand, this same advantage will often lead to a welfare ranking that is not very discerning and that becomes less so as the number of choice observations increases. In such circumstances, many pairs of alternatives x and y are not comparable under P^* (hereinafter denoted xN^*y). To handle this criticism, B&R propose to “prune” \mathcal{G}^* by using non-choice data to delete “suspect” GCSs. This refinement allows to identify a *welfare-relevant domain* $\mathcal{G} \subseteq \mathcal{G}^*$, that consists of the GCSs from which the observer takes normative guidance. The next two approaches I will discuss can be represented as criteria for pruning \mathcal{G}^* , and I will illustrate their welfare inferences doing so. For now, we can represent the introductory examples and the context-independent

¹⁰An observer can be a social planner such as a government official or a CEO (as in B&R), but also an expert that wishes to give an advice to an agent or a mediator that wishes to facilitate a contract between parties. Observers are assumed to be benevolent and impartial.

¹¹B&R make the auxiliary assumptions that (i) for all subsets $A \in X$ there is some ancillary condition d for which $(A, d) \in \mathcal{G}^*$ and that (ii) $C(A, d)$ is non-empty for all $(A, d) \in \mathcal{G}^*$ (p. 56). I comment on the implications of dropping (i) in section 4.

criterion’s welfare inferences as follows:¹²

Table 1: The snacks choice (context-independent criterion)

Generalized choice situation, (A, d)	Chosen alternative, $C(A, d)$	Welfare-relevant domain, \mathcal{G}
$(\{apple, snickers\}, advance\ choice)$	<i>apple</i>	$(A, d_1) \in \mathcal{G}$
$(\{apple, snickers\}, immediate\ choice)$	<i>snickers</i>	$(A, d_2) \in \mathcal{G}$

Welfare inference: *apple* N^* *snickers*.

Table 2: The smoking choice (context-independent criterion)

Generalized choice situation, (A, d)	Chosen alternative, $C(A, d)$	Welfare-relevant domain, \mathcal{G}
$(\{smoking, not\ smoking\}, abstinence)$	<i>smoking</i>	$(A, d_1) \in \mathcal{G}$
$(\{smoking, not\ smoking\}, not\ abstinence)$	<i>smoking</i>	$(A, d_2) \in \mathcal{G}$

Welfare inference: *smoking* P^* *not smoking*

In the *snacks choice* example (Table 1), the *apple* and the *snickers* are not comparable according to the context-independent criterion because the chosen alternative in the advance choice condition (d_1) is contrary to the one chosen in the immediate choice condition (d_2). In other words, the context-independent criterion is agnostic about which snack is welfare superior because there are conflicting choice patterns. In the *smoking choice* example (Table 2), the criterion is discerning since *smoking* is consistently chosen over *not smoking*. Hence, the former alternative is said to improve upon the latter. In both cases there is no pruning of \mathcal{G}^* (i.e., all GCSs belong to the welfare-relevant domain \mathcal{G}).

3. Reason-Based Choices

Even when confronted with the evidence that observed behaviour is often “inconsistent”, economists usually take the satisfaction of a given stable and context-independent preference as the benchmark for welfare analysis (e.g. Koszegi and Rabin 2007; Rubinstein and Salant 2012; Apesteguia and Ballester 2015). One approach among these, that underlies many recent economic models, is to assume that only choices that are made after rational deliberation (*reason-based choices*) are reliable proxies of welfare.¹³

¹²B&R’s preferred method to prune \mathcal{G}^* is to rely on evidence gathered from psychology, neuroscience, and neuroeconomics on informational processing failures, such as on the incorrect use of information, lack of attention, or naive forecasting (pp. 83-5). Their refined criterion is then similar to the one I will turn next (see also Infante et al. 2016). Whenever referring to the context-independent criterion I mean the more circumscribed approach that does not take this refinement into account.

¹³What exactly counts as rational deliberation is contestable and varies according to the model. In this section I present two different but for the most part compatible conceptions of rational deliberation derived from psychology and philosophy. See Infante et al. (2016) for a critical review of this general approach.

A prominent example is given by “dual-system” models recently popularized by Kahneman (2011) (see Alós-Ferrer and Strack 2014 for a short review of different theories). According to this view human psychology can be divided into two systems or modes of thought: one fast, effortless, and automatic (System 1), and another slow, effortful, and controlled (System 2). In economics these models have been used, among other things, to represent intrapersonal conflict between present and future preferences (e.g. Thaler and Shefrin 1981; Bernheim and Rangel 2004; Fudenberg and Levine 2006, 2012). For instance, Bernheim and Rangel (2004) build a model to study addictive behaviour in which an agent alternates between a “hot mode” and a “cold mode”. Whenever “cued” towards the hot mode the agent always takes an addictive behaviour “irrespective of underlying preferences”, while in the cold mode she “considers all alternatives and contemplates all consequences” and selects her most preferred alternative (p. 1559). They assume that agents maximize a context-independent and stable preference relation on their cold mode and that choices taken under the hot mode are “mistakes”. Reason-based choices are assumed to be reliable (and consistent) proxies of welfare.

A criterion based on rational deliberation has also found support in a recent influential book by Hausman (2012). Hausman (2012) aims to describe how the concepts of *preference*, *value*, *choice*, and *welfare* are and ought to be used in economics. Hausman argues that the concept of a single preference, as employed in neoclassical economics, is and ought to be a *total subjective comparative evaluation* (TSCE). It is *comparative* in the sense that people prefer one state of affairs to another. It is *subjective* in the sense that the comparison is made from a first-person perspective. And it is *total* in the sense that it is a comparison that takes into account everything that the agent considers to be relevant for choice. It seems clear that preferences are comparative and subjective, and nowadays, many economists would agree that the concept of preference, as used in neoclassical economics, is often an all-things-considered ranking of alternatives (see also Baigent 1995, p. 92).

But according to Hausman (2012) a preference is and ought to be also an *evaluation*, in the sense that it is the result of a rational deliberation about what agents have most reason to do. Hausman (2012) argues that a preference should be seen as a reason-based evaluation rather than a judgement, rather than an expression of taste, and rather than a feeling, because judgements do not by themselves motivate action, tastes do not exhaust the considerations relevant to choice, and feelings by themselves do not provide reasons for action (see also Hausman 2013, p. 219). This means that an agent’s choice that is not based on a rational deliberation about what she has most reason to do is not considered to reveal a preference. It follows that such a choice, according to Hausman’s view, cannot be used as a proxy of welfare. Only reason-based choices are *potential* proxies of welfare.¹⁴

¹⁴Hausman (2012) argues that preferences are reliable proxies of welfare only when these are self-interested, informed, and competently considered (see also Hausman and McPherson 2009 and Hausman 2016). Infante et al. (2016) argue that in this general approach — here referred to as reason-based — most authors, including

B&R’s framework can be used to represent possible welfare inferences of this approach. Evidence concerning internal deliberation *before choosing*, if rational or not, can be used to delete certain GCSs (i.e., to prune \mathcal{G}^* and identify the welfare-relevant domain $\mathcal{G} \subseteq \mathcal{G}^*$). Take the introductory examples, and suppose that we have gathered evidence that both immediate consumption and abstinence trigger impulsive behaviour (or System 1). Conversely, suppose that evidence strongly suggests that both being far from consumption and abstinence encourage reason-based deliberation (or System 2). Since it is difficult to determine if deliberation is rational or not for specific choices of individual agents, observers often rely on indirect data like this. Then:

Table 3: The snacks choice (reason-based criterion)

Generalized choice situation, (A, d)	Chosen alternative, $C(A, d)$	Welfare-relevant domain, \mathcal{G}
$(\{apple, snickers\}, advance\ choice)$	<i>apple</i>	$(A, d_1) \in \mathcal{G}$
$(\{apple, snickers\}, immediate\ choice)$	<i>snickers</i>	$(A, d_2) \notin \mathcal{G}$
Welfare inference: <i>apple P* snickers.</i>		

Table 4: The smoking choice (reason-based criterion)

Generalized choice situation, (A, d)	Chosen alternative, $C(A, d)$	Welfare-relevant domain, \mathcal{G}
$(\{smoking, not\ smoking\}, abstinence)$	<i>smoking</i>	$(A, d_1) \notin \mathcal{G}$
$(\{smoking, not\ smoking\}, not\ abstinence)$	<i>smoking</i>	$(A, d_2) \in \mathcal{G}$
Welfare inference: <i>smoking P* not smoking</i>		

In the *snacks choice* example (Table 3), this criterion leads to a welfare ranking that is *finer* (or more discerning) than the one derived from the context-independent criterion. According to the reason-based criterion, it is possible to infer that the *apple* is welfare superior to the *snickers* because the choice on the advance condition was likely to be reason-based (System 2) while the choice on the immediate condition was not (System 1). In the *smoking choice* example (Table 4), the inference is the same as before even if the rationale behind it is different. *Smoking* is said to be better than *not smoking*, not because Norah is a consistent smoker, but because she has chosen to do so after rational deliberation.

Hausman (2012), are assuming that individuals are endowed with an *inner rational agent*: “[H]uman psychology is represented as a set of forces which affects behaviour by *interfering with* rational choice [and] rational choice itself [is] represented by the error-free reasoning of [an] inner agent” (pp. 14-5, italics in original). Infante et al. (2016) argue that this model is not tenable because even if one accepts the dual-system theory as a useful way of organising ideas about human psychology, one is not entitled to assume that reason-based choices are consistent across context and time (pp. 14-5). This assumption is shared by many of the “model-based approaches” that try to identify the cognitive process that underlies choice through standard revealed preference methods, and that under the additional assumption that the model is true, claim to identify which choices are welfare-relevant and which ones are mistakes (see Manzini and Mariotti 2014, p. 347).

4. Confirmed Choices

I have distinguished two prominent welfare criteria adopted in behavioural economics that equate some (as opposed to all) choices with welfare. In both cases an objective standard of rationality is used to determine which choices are welfare-relevant. I wish now to propose an alternative criterion that relies on individual subjective judgements instead. As I will try to argue below, using this criterion is conducive to the satisfaction of individual welfare-enhancing preferences (section 4.1), to a more accurate link to subjective well-being than the previous criteria (section 4.2), and to the respect of individual sovereignty (section 4.3).

I propose to identify choices that are reliable proxies of welfare according to an agent's confirmation of those choices. Consider a discrete time horizon $\mathcal{T} = \{0, \dots, T\}$ and assume without loss of generality that GCSs are ordered in time from 1 to $T - 1$. An agent is said to *confirm* her own choice made at period $t \in \mathcal{T}$ if, when reflecting at T upon facing that choice again, she would not want herself to choose differently at T . This means that an agent confirms her own choice if when reflecting at a “distance” from her decision she would not want to change her behaviour.¹⁵ For simplicity, I assume that the observer makes their welfare evaluation marginally after period T .¹⁶

Asking Norah if she confirms or not her choice of x (*watch a film*) over y (*studying*) corresponds to ask her to perform the following thought experiment: “From where I stand now, would I want myself to choose x over y again?” Then, if Norah responds that she would want herself to choose it again, her actual choice to *watch a film* instead of *studying* seems to have revealed a ranking between x and y that Norah (and not an observer) deems relevant for her welfare at the time of the welfare evaluation. I call such choice a ***confirmed choice***.

Conversely, Norah may not confirm her choice. Upon reflection, it is possible that Norah would want herself to behave differently if faced with the same choice again. I call this kind of choice an ***unconfirmed choice***. There are several possible reasons for an agent *not confirm* one of her choices. For example, (i) Norah may wish to have chosen differently (e.g. she regrets her choice to *watch a film*), (ii) she may change her mind about the worth of the alternatives (e.g. Norah forms a new ideal to be a good student), (iii) she may acquire additional information (e.g. Norah learns that she has an unexpected exam), or (iv) she may change her intentions regarding the alternatives (e.g. she loses her desire to watch films). This means that Norah may not confirm her choice even though she does not regret it and she has not learned additional

¹⁵Noor (2011, p. 602) argues that a “normative preference is revealed when the agent is distanced from the consequences of his choices” and assumes that this preference is best identified the later the consequences are realized. Contrary to Noor (2011) I focus on the distance to the decision and make no assumption concerning optimal distance.

¹⁶A similar analysis can be made with confirmations at period 0 and welfare evaluation marginally after period 0. In that case choices are expected/predicted (as opposed to made/observed). Confirmations at T are preferable for *ex post* policy (welfare) analysis, which is particularly useful if it is possible to experiment before implementing the intended policy. *Ex ante* policy analysis is still possible with confirmations at T , since the consequences of the agents' choice(s) may realize after T .

information.¹⁷

We can use B&R’s framework to represent possible welfare inferences of this criterion. In this case, “who” prunes \mathcal{G}^* is no longer the observer but the agent herself. The distinction between confirmed and unconfirmed choices can be revealed through individuals’ verbal judgements of their choices: “From where you stand now, what would you want yourself to choose?” This data is used in an “auxiliary role”, since the justification for implementing an alternative rather than another is still primarily linked to the agent’s choices: its role is to help the observer to make an educated guess about the agent’s *judgements* about her choices.¹⁸ Self-authentication provides a non-arbitrary justification to include or exclude choices from the welfare-relevant domain, precisely because it is based on the agent’s own judgements about her choices: “if we are measuring the extent to which a life is valuable for the person living it, then it seems that the criteria for evaluation must be those of the agent herself” (Noggle 1999, p. 303).

Take again the introductory examples. Recall that Norah would want herself to choose *not smoking* if faced with the same choices again. For the sake of illustration, suppose that Norah confirms her choice of *snickers* over *apple* but not her choice of *apple* over *snickers* (say because she is usually low on sugar). Then:

Table 5: The snacks choice (confirmed criterion)

Generalized choice situation, (A, d)	Chosen alternative, $C(A, d)$	Welfare-relevant domain, \mathcal{G}
$(\{apple, snickers\}, advance\ choice)$	<i>apple</i>	$(A, d_1) \notin \mathcal{G}$
$(\{apple, snickers\}, immediate\ choice)$	<i>snickers</i>	$(A, d_2) \in \mathcal{G}$

Welfare inference: *snickers* P^* *apple*.

Table 6: The smoking choice (confirmed criterion)

Generalized choice situation, (A, d)	Chosen alternative, $C(A, d)$	Welfare-relevant domain, \mathcal{G}
$(\{smoking, not\ smoking\}, abstinence)$	<i>smoking</i>	$(A, d_1) \notin \mathcal{G}$
$(\{smoking, not\ smoking\}, not\ abstinence)$	<i>smoking</i>	$(A, d_2) \notin \mathcal{G}$

Welfare inference: *smoking* N^* *not smoking*

¹⁷In general the nature of judgements can be multiple (e.g. concerning reasons for action, subjective well-being, all-things-considered), and my arguments do not hinge upon a particular assumption. My arguments neither depend upon judgements being action-guiding or not, especially because the judgements I consider are asynchronous with respect to the moment of choice. I also impose no requirement on the kind of reflection that precedes confirmation. On the one hand, this favours parsimony and generality. On the other hand, it is possible that Norah does not give serious consideration to her reflection. Similarly, Norah’s judgements may be affected by frames, cognitive bias, and other sources of context-dependency that can limit the reliability of the data collected. I discuss these latter issues in section 5.

¹⁸Some economists will resist this approach because its empirical content relies in part upon non-choice data. There exist by now, however, some controlled ways to partly circumvent the usual concerns about inattention and deception. In experiments and experimental surveys it is possible to record, among other things, the duration taken by subjects to give an answer in order to exclude speedy answers that cannot be the result of honest attentive answers. See Manzini and Mariotti (2014, p. 344) for an argument in favour of using non-choice data for a similar purpose.

The *snacks choice* example illustrates that a welfare ranking based on confirmed choices is not necessarily aligned with some objective view of the good or prudent reason for action (Table 5). It also shows that pruning \mathcal{G}^* with confirmations can lead to a welfare ranking that is more discerning than one based on context-independent choices. The *smoking choice* example illustrates that an observer using this criterion is prudent when contemplating unconfirmed choices (Table 6). In those cases, the observer is agnostic about which alternative is welfare superior. In this example, if smoking is an addiction and/or an irresistible desire for Norah, the criterion can be said to be sensitive to the fact that smoking is against Norah’s will.

Formally, note that P^* is not necessarily acyclic under a domain \mathcal{G} pruned with an agent’s confirmations. This may be problematic since without (a guarantee of) acyclicity it may not be possible to identify welfare optima for some choice situations and/or unambiguous welfare improvements. Note, however, that this limitation is shared by most refinements of B&R’s framework. In fact, B&R’s preferred welfare ranking P^* is only (guaranteed to be) acyclic when one observes the agent choosing from at least all two-element and three-element subsets of X , which in practice rarely occurs. In section 4.1 I will argue that even though confirmations do not guarantee acyclic welfare rankings they have the potential to increase their likelihood in restricted domains.

These claims illustrate some of the merits and limitations of B&R’s framework and of pruning \mathcal{G}^* .¹⁹ It is however possible to use confirmations on other frameworks and applications. For example, Chambers and Hayashi (2012) propose a mapping from stochastic choice (the empirical frequency of chosen alternatives for any possible choice situation) into a transitive and complete welfare binary relation based on a few axioms describing the consistency of welfare across data and weights on every (*chosen alternative, choice situation*) pair. If interested in having a complete and transitive welfare ranking of alternatives, an observer can use empirical data on confirmations to select “reasonable” weights for the different (*chosen alternative, choice situation*) pairs.

Later in the paper I consider policy implications outside B&R’s framework that rely on the distinction between confirmed and unconfirmed choices (section 5). For now I turn to three complementary arguments that support confirmed choices as a reliable (though fallible) proxy of welfare, or, at least, that support the necessity of using confirmations to make reliable welfare comparisons from choice behaviour.

¹⁹Pruning \mathcal{G}^* is used to illustrate, in a simple manner, the inferences of the different criteria. I make no claim of its methodological superiority (see Bernheim 2009; cf. Manzini and Mariotti 2014). For example, pruning \mathcal{G}^* may be seen as overly prudent for domains that include three-element or larger subsets of X . An observer may wish to contemplate the possibility that an agent may choose x over y and z ($C(\{x, y, z\}, d) = x$), and “confirm” the choice of x over y but not that of x over z . By deleting $(\{x, y, z\}, d)$ from \mathcal{G}^* the observer would lose relevant information. Remark that if an agent’s confirmations are available and they are considered “the best way to disrespect choice” (Manzini and Mariotti 2014, p. 347), an observer that wants to respect the agent’s interest does not need an underlying model of decision making to elicit welfare in the sample observed. However, an observer may use an underlying model of decision making and revealed preference techniques to try to understand which type of choices the agent does and does not confirm in order to make out of sample predictions.

4.1 The Argument From Preference Satisfaction

In neoclassical economics preference satisfaction is one of the main views of individual welfare.²⁰ Among the many usages of the term “preference”, it is often used to refer to an all-things-considered ranking of alternatives that is not necessarily revealed through choices (e.g. Baigent 1995; Hausman 2012), or, alternatively, to refer to the choice-ranking of alternatives (e.g. Harsanyi 1997; Sugden 2018).

In what follows I start from the premise that all choices reveal a preference-ranking of alternatives (a *revealed preference*), but that only some choices reveal preferences that are aligned with the agents’ interests (*welfare-enhancing preferences*; see Beshears et al. 2008 for a similar distinction). To identify welfare-enhancing preferences it is essential, I claim, to make a distinction between the preferences that agents want satisfied and the preferences that agents do not want satisfied (hereinafter *willed* and *non-willed* preferences respectively). As pointed by Sagoff (1986, p. 303), it is false that each person wishes her own preferences to be satisfied: “A person wishes his preferences satisfied at the moment he has them, but he often changes his mind, regrets they were satisfied, or is grateful they were not.”

With these premises in mind, I argue that confirmations are necessary in order to identify which choices reveal welfare-enhancing preferences. Let me illustrate with a comparison to reason-based and context-independent choices. Take first the example of a choice that results from an informed and competent (rational) deliberation. According to the reason-based criterion this choice reveals a welfare-enhancing preference. However, a reason-based choice may reveal a non-willed preference. The fact that one has deliberated about reasons to act before choosing does not exclude the possibility of not wanting her revealed preference to be satisfied at a different moment in time. For example, suppose that Norah faces a single-shot decision between purchasing standard or premium travel insurance, and assume that she chooses standard after a slow considered deliberation. Confronted with the unexpected responsibility she felt for the choice, Norah may wish she had chosen premium instead (see Botti and McGill 2006 for the effect of perceived responsibility on post-choice satisfaction). Her choice has revealed a non-willed preference, and it is likely that she would want to behave differently when reflecting upon facing that choice again. Then, while a criterion based on confirmed choices is likely to be sensible to the fact that Norah does not want her preference satisfied, the reason-based criterion (and the context-independent criterion alike) is not.

Conversely, a choice may reveal a willed preference even though it is not based on a rational deliberation. Take the example of the large number of choices made out of habit. Many of these choices can be said to be “made with good reason although not deliberated” (Broome

²⁰The claim that a person’s well-being consists of the satisfaction of her desires (or some subset of them) is also influential in philosophy. Though some authors defend that well-being consists of the satisfaction of any desire (e.g. Lemaire 2016), most define some condition(s) to exclude some desires (e.g. Sidgwick 1907; Brandt 1979; Lewis 1989; Rosati 1995; Noggle 1999). I discuss fully-informed accounts of well-being in section 5.

1978, p. 326). It is not clear, then, why to disregard the preference-rankings revealed by these choices as welfare-enhancing. The reason-based criterion is in this case overly restrictive as it is likely to unduly reduce the number of choices in the welfare-relevant domain. Taken together, these arguments suggest that a reason-based criterion is neither necessary nor sufficient for the satisfaction of welfare-enhancing preferences.

The same holds for context-independent choices. A context-independent criterion seems again overly restrictive, as it is likely to discard choices that reveal preferences that agents want satisfied. A notable example is *changes of preference* (when understood as changes of mind or intention). For instance, suppose that Norah used to choose *pork-chops* when *veggie-roast* was available in her local restaurant, but that now she chooses *veggie-roast* instead because she has formed a new ideal in favour of no animal suffering. According to the context-independent criterion *pork-chops* and *veggie-roast* are not comparable in terms of welfare. However, it seems that welfare rankings should ignore past choices that are no longer judged of worth or important (see Parfit 1984, ch. 8 for a related comment). Conversely, context-independent choices may reveal preferences that an agent does not want satisfied. Norah's choice between travel insurances illustrates this case.

A confirmed criterion is, by construction, sensible to the fact that some choices reveal willed preferences while others reveal non-willed preferences. To see this, consider the seminal experiments on individual decision making under risk that show that many subjects do not behave according to the postulates of expected utility theory (e.g. Tversky and Kahneman 1981; Starmer and Sugden 1989). Since then, several experiments have shown that many subjects revise their decisions in the direction of theoretical predictions when repeating them, either learning or not the consequences of their actions (e.g. van de Kuilen and Wakker 2006; van de Kuilen 2009; Birnbaum and Schmidt 2015; cf. Cubitt et al. 2001). In such cases, a first approximation is to infer that subjects that revise their decisions have revealed, in their first choices, preferences that they do not want satisfied at the end of the experiment. This inference may be however mistaken (e.g. some subjects may have a preference for randomization). The "guess" about subjects' judgements becomes more "educated" if subjects are given the chance to evaluate their past decisions. That is the type of empirical data that a confirmed criterion asks for, and an observer acting on that information would be able to exclude choices from the welfare-relevant domain that have revealed preferences that the agents themselves would not want satisfied.

The last example also illustrates that by encouraging asynchronous reflection, a confirmed criterion has the potential to increase the likelihood of finer and acyclic welfare rankings (see also Gilboa 2010, p. 7 and Kahneman and Tversky 1979, p. 277). Similarly, by encouraging agents to reflect upon their behaviour an observer is more likely, I conjecture, to delete choices from the welfare-relevant domain that are the result of arational behaviours than other choices. If so, this criterion can be conducive to the satisfaction of willed and sometimes acyclic preferences.

These arguments, if correct, suggest that confirmations are necessary in order to identify which choices reveal welfare-enhancing preferences. I do not wish to claim that they are sufficient. As with choices and preferences, a person can change her mind about her judgements. For example, Norah may form an even stronger desire to smoke and consider that the proxy of welfare based on confirmed choices (Table 6) is misaligned with her new interests. This is less problematic at the moment that the welfare evaluation is made because the observer is following Norah’s own judgement at that point in time. However, it shows that confirmed choices can be “unconfirmed” in the future by the agent herself. Confirmations can also be affected by adaptation to one’s opportunities, manipulation, context-dependency, and other phenomena that most observers regard (I believe rightly so) as worrisome for welfare analysis based on subjective data (I return to these concerns in section 5). The argument of this section is rather that data on confirmations is necessary for an observer to identify, “here and now”, which choices reveal welfare-enhancing preferences as judged by the agent herself.

4.2 The Argument From Subjective Well-being

There has been a renewed interest in the notion and measurement of subjective well-being (hereinafter SWB; see e.g. Ryff 1989; Kahneman et al. 1997; Kahneman and Riis 2005; Benjamin et al. 2014). While the term SWB is often associated to happiness or life satisfaction, there is a growing recognition that SWB is multidimensional and governed by some “fundamental aspects” (e.g. Adler and Dolan 2008; Kahneman and Deaton 2010; Benjamin et al. 2014). In this section, I wish to argue that confirmed choices are a more accurate proxy of (multidimensional) SWB than context-independent and reason-based choices.

One argument in favour of this claim derives from the previous section: higher preference satisfaction is likely to be associated with higher SWB (see e.g. Benjamin et al. 2012). This seems to be particularly the case when preference satisfaction is restricted to willed preferences. Then a corollary of the previous argument, if correct, is that confirmed choices are likely to be a better proxy of SWB than context-independent and reason-based choices as far as preference satisfaction is concerned. I wish now to argue that confirmations trace other fundamental aspects of SWB such as living according to personal values, being *who* one wants to be, avoiding self-control costs and some negative emotions.

Consider the case of living according to personal values and being who one wants to be. These two aspects are part of what is often called the *eudaimonic* measures of SWB, linked to the individual interest in having a meaningful, valuable, worthwhile life (e.g. Ryff 1989; Kirman and Teschl 2006). Benjamin et al.’s (2014) evidence supports the high relative marginal utilities of these aspects of well-being on overall SWB. In a series of hypothetical-choice scenarios, they asked a large U.S. adult population to make trade-offs between different aspects of SWB, two at a time, derived from a comprehensive list of more than 100 aspects. According to their estimates

of the relative weight of each aspect on overall SWB, “You being a good, moral person living according to your personal values” is ranked 4th and “You ‘being the person you want to be’ ” is ranked 22nd on the *personal* aspects of SWB (p. 2715). Confirmations bring information, not captured in individuals’ choices, about what people value and care about (see Hirschman 1984 and Lewis 1989, p. 115 for related arguments). For instance, Norah’s confirmation of the choice of *veggie-roast* over *pork-chops* in the example above is an indication that that choice is aligned with her personal values and who she wants to be.

Turning to self-control, there is by now considerable evidence that some individuals are willing to self-impose commitments to limit self-control costs and avoid negative consequences (e.g. Ashraf et al. 2006; Augenblick et al. 2015; Bonein and Denant-Boèmont 2015). For example, in a recent experiment that aims to test dynamically inconsistent time preferences in effort, 59% of subjects demand a commitment to their initial effort allocation choice at price \$0 (Augenblick et al. 2015). Augenblick et al. (2015) also find that a bias towards the present is predictive of the demand of commitment. Together, these findings suggest that some subjects care about potential costs associated with a lack of self-control and that some with higher self-control problems are self-aware of their limited self-control. At the same time, it is possible that some subjects that are aware of their limited self-control confirm their own choices. Then, the distinction between confirmed and unconfirmed choices seems relevant if an observer wishes to make reliable welfare comparisons when “tempting” alternatives, like smoking and procrastination, are available against “non-tempting” ones.

Consider now negative emotions. According to some authors, negative emotions are an important aspect of SWB that should be traced by policymakers (e.g. Deaton et al. 2011). On the one hand, negative emotions like stress, anger, and anxiety are likely to escape most welfare criteria that are primarily based on choice behaviour. On the other hand, negative emotions like regret (“I wish I had not done that”) and melancholy (“I wish I had”) can be revealed through confirmation of behaviour (see Read 2006 for this distinction). In both cases, an agent is likely to want herself to behave differently when reflecting upon repeating a choice that she regrets or is melancholic about. This claim extends to cases in which the context-independent and reason-based criteria would fail to trace these emotions. For example, in single-shot decisions, regret and melancholy will not translate into behaviour. Then, both the context-independent and reason-based criteria will fail to recognize that the agent “wishes she had not chosen that alternative” or that she “wishes she had chosen another alternative” while a confirmed criterion will not.

Notwithstanding, confirmed choices do not trace some important aspects of individuals’ interests such as positive emotions, some negative emotions, and aspirations that are unrelated to choice behaviour. This can be captured by SWB measures that incorporate experiential measures and/or elicit agents’ goals. A criterion based on confirmed choices also fails to provide any

information on the quality and intensity of individuals' experiences and their memories of those (see Kahneman and Riis 2005). Remark that these concerns are, however, shared by the criteria based on context-independent and reason-based choices.

Hence, though choices and confirmations are certainly not sufficient to measure SWB, together they trace aspects of well-being, such as personal values and regret, that may escape measures that rely on other choices, experience, neural activity, or stated preferences.

4.3 The Argument From Individual Sovereignty

My last argument concerns the respect of individual sovereignty. In economics the respect of individual sovereignty, often referred to as consumer sovereignty, has been traditionally seen as a grounding principle: individuals' own judgements are treated as decisive in assessing the relative welfare associated with different alternatives or states of affairs. In most cases consumer sovereignty is associated with the respect of individual choices or preferences. For example, Sugden (2004, 2018) has recently argued in favour of a robust concept of consumer sovereignty that attaches value to an agent's opportunities to act as she wants.

Individual sovereignty is important for welfare inferences at least for two reasons. First, in many circumstances "individuals are better placed than third parties to determine what choices enhance their own well-being" (Mill 1859 [2010], ch. 3-4; cf. Sunstein and Thaler 2003). Waldfogel (2005) provides some evidence that this is the case for choices among current consumption goods. Using data from the Christmas/Hanukkah gift-giving season, Waldfogel (2005) finds that individuals value their own purchases at an average of 18% more, per dollar spent, than they value items they receive as gifts from their friends and family (excluding sentimental value). The respect of individual sovereignty, in this case, can be instrumental to enhance agents' well-being and to have more reliable proxies of welfare in some contexts of choice. Second, the reasons for deferring to an agent's judgement go beyond her reliability as a judge (Velleman 1999, p. 608; see also Mill 1859 [2010]). Being independent, capable of self-determination, and other autonomy-related concerns are part of the subjective interest of many individuals (see Benjamin et al. 2014 for empirical data in support of this claim). They are procedural concerns that have an intrinsic value for individual welfare, and that are satisfied through the respect of individual sovereignty.

In this section, I wish to argue that the respect of individual choices is not enough to respect individual sovereignty. Not surprisingly, I will argue that a broader notion that concerns agents' choices and their judgements about their choices is required.²¹

²¹See Decancq et al. 2015 (2015, p. 1083) for a related principle of individual sovereignty that requires the respect of "informed judgements" that are not necessarily revealed in agents' choice behaviour. Sugden (2004) argues that respecting second-order preferences is opposed to individual sovereignty because second-order preferences locate normative authority "not in the day-to-day decisions that individuals make as economic actors, but in each person's supposed higher moral self" (p. 1017). A robust concept of individual sovereignty, according to his view, "should not need to invoke such a moralized account of preference". It is worth noting that part of this criticism is at odds with Frankfurt's (1971, fn. 6) view that second-order attitudes are not necessarily moral,

Take the *smoking choice* example again. Why should not Norah's judgement about her choice be important when considering her sovereignty? Why is her revealed preference more important than her judgement of her choice in this respect? I propose that both are relevant and that it is possible to respect Norah's choice sovereignty even when taking her (confirmation) judgements into consideration. Indeed, it is even possible to reinforce Norah's individual sovereignty when a broader notion of sovereignty is adopted. Let me illustrate.

Consider a social planner that is interested in providing opportunities to Norah. If they want to respect her individual sovereignty and the value of opportunity they should not prohibit smoking (Sugden 2004). This would indeed reduce her opportunities and be against her actual choices. But if a planner wants to respect Norah's sovereignty they should not only not prohibit smoking but also implement some policy that would help her to follow the behaviour that, according to her judgement about her behaviour, she would want herself to follow (*not smoking*). Providing Norah with free consultations with a specialized doctor would be such a policy. By doing so, the planner would respect Norah's judgement and her opportunity to lead the life that according to her judgement she seems to value. Without such policy, Norah has the *negative freedom* to go to a consultation (it suffices that it is available on the market), but she may not have the *positive freedom* to do so, for, say, budget constraints.²² Since the two policies, non-prohibition and free consultations, are not mutually exclusive, the planner would respect Norah's sovereignty in terms of her choices and (confirmation) judgements. This would reinforce her individual sovereignty.

In light of these arguments, confirmation of choices seems necessary in order to fully respect individual sovereignty. Thus respecting confirmed choices, if this argument holds, is likely to be conducive to the respect and reinforcement of individual sovereignty. On the contrary, reason-based choices rely on an objective view of rationality that may not be endorsed by agents and that is likely to violate individual sovereignty. Similar concerns hold for context-independent choices. This last claim may be surprising since B&R's main justification to respect choices is the respect for agents' self-determination (see also Bernheim 2009, pp. 291-2). However, the respect for self-determination seems to provide neither a rationale to rely exclusively on choices nor a rationale for why to select context-independent choices (rather than others) when conflicting choice patterns are present. An observer that assumes that context-independent choices are welfare-enhancing may be making a judgement that is contrary to the agent's own judgement. They could then fail to respect her sovereignty and, as the previous example illustrates, fail to provide the opportunities that she values.

but instead the evaluations, no matter their origin, that an individual makes about the different desires that she holds. Similarly, confirmation does not necessarily rely on a higher moral self. For instance, a friend of Norah may confirm her choice of not eating meat, not because she finds it immoral to eat meat, but because she cares about her health or simply because not eating meat is economically-wise.

²²In the most famous formulation of the distinction between negative and positive freedoms, due to Berlin (1969), the former is interpreted as the freedom from constraints that are imposed by others and the latter is interpreted as the ability to lead a life in an autonomous and reasoned fashion (p. 8).

5. Informed Choices, Libertarian Paternalism, and Policy Implications

These three arguments, when taken together, provide an additional liberal reason to satisfy choices even when these seem contrary to reason. They suggest, I claim, that confirmed choices may be a reliable (though fallible) proxy of welfare. However, as I have previously suggested, the cases in which this proxy is fallible are not negligible. Familiar objections to subjective welfare criteria such as incorrect and/or incomplete information (e.g. Harsanyi 1997), context-dependency of subjective attitudes (e.g. Sugden 2004), or adaptation to existing opportunities (e.g. Elster 1982) still have force under the view endorsed here.

Partly for these reasons many authors endorse fully-informed hypothetical preferences when considering individual welfare: a person's well-being consists of the satisfaction of the desires that she would have if she had all relevant information and, for some authors, made full rational use of this information (e.g. Sidgwick 1907; Brandt 1979; Arneson 1990; Harsanyi 1997).²³

This principle has recently been invoked as a normative justification for *libertarian paternalistic* interventions, that wish to steer "people's choices in welfare-promoting directions without eliminating freedom of choice" (Sunstein and Thaler 2003, p. 1159) (hereinafter S&T).²⁴ Individuals are said to take decisions that do not promote their own welfare if these are "decisions that they would change if they had complete information, unlimited cognitive abilities, and no lack of self-control" (S&T, p. 1162). S&T then argue that "private and public institutions" shall steer people in the direction of a hypothetical (fully-rational) *informed choice*. Doing so will increase, they argue, the welfare of targeted individuals.

Identifying hypothetical informed choices (or preferences) is, however, a difficult exercise (see e.g. Cowen 1993; Sobel 1994, 2009; Rosati 1995; Noggle 1999; Fumagalli 2016). It is not necessarily straightforward to define what is complete information, unlimited cognitive abilities, or no lack of self-control. For instance, how much and what type of information is necessary for one to be fully-informed about some topic? Rosati (1995) and Noggle (1999) argue that this reasoning leads to an infinite regression into the view that a person's well-being is tied to the desires that she would have if she were an "ideal observer". This ideal observer would not only be fully informed but would also have to have experienced all available options, had education, training, and so on. But even if such an ideal observer is not required, hypothetical informed

²³Others, like Sugden (2004), conclude that normative economics should not use subjective information but instead rely on opportunities to act. According to Sugden (2004), as soon as we acknowledge that preferences change according to "trivial changes in viewpoint or context", then using preference satisfaction (either of first- or second-order) for normative analysis is not possible (p. 1016). I agree that if preferences are invariably contingent then normative analysis is better done without taking preferences into consideration. Still, there seem to remain cases in which preferences and judgements are (temporarily) stable or change for predictable and reasonable reasons. There is also by now considerable evidence "that provides constraints on the types of preferences that one can reasonably invoke" (see Fehr and Hoff 2011, pp. 398-400). Despite these arguments, I side with Sugden (2004) in that objective normative measures are necessary, but I claim that these should complement, rather than substitute, subjective measures.

²⁴See also Thaler and Sunstein (2003, 2008). See Camerer et al. (2003) for a similar approach. The common approach of these papers, often called soft paternalism, has been well received by many behavioural economists. See Sugden (2008), Qizilbash (2012), and Fumagalli (2016) for critical reviews.

preferences do not exist now as they pertain to the preferences with information and cognitive abilities that the agent does not have:

“The preferences of perfectly informed individuals are not always relevant for imperfectly informed choice. By considering perfectly informed preferences, we are hypothetically changing an individual’s human capital endowment. What an individual would want with a different human capital endowment cannot necessarily be extrapolated usefully into information about what improves the welfare of an individual now.” (Cowen 1993, p. 262)

Further, relying on hypothetical informed choices is likely to be contrary to the individual sovereignty of the actual agents. It is neither clear that relying on hypothetical informed choices improves upon confirmed choices in terms of preference satisfaction and subjective well-being. Fumagalli (2016) has recently pointed several epistemic and evidential concerns that suggest that “the new paternalists typically lack the information required to design and implement welfare-enhancing paternalistic interventions” (p. 471). But even if one disregards the evidential and epistemic difficulties faced by new paternalists, a fully-informed person may still not endorse the satisfaction of a given preference or fail to desire to make a choice that she confirms or believes would be good for her (see also Noggle 1999, pp. 305-6). This means that even if one supports a notion of well-being as the satisfaction of hypothetical informed preferences (or hypothetical informed choices as proxies of welfare), confirmation of those preferences (choices) is still necessary.

Nonetheless, the concerns about lack of information, bounded rationality, among other underlying reasons for “ill-formed” subjective attitudes highlighted by S&T and others are significant. I agree with S&T that it is false that “almost all people, almost all of the time, make choices that are in their best interest or at the very least are better, *by their own lights*, than the choices that would be made by third parties” (p. 1163, italics added). This statement, however, does not justify libertarian paternalistic interventions that rely on some hypothetical view of well-being, particularly if “by their own lights” is interpreted, as it is implicit in S&T, as the judgements of hypothetical fully-informed rational agents.

The arguments exposed so far point towards a viable alternative to soft paternalism: *steering people in the direction they can form judgements about their behaviour*.²⁵ Several policies are in line with this perspective. One example is providing opportunities for individuals to engage in deliberative decision-making (see also Voorhoeve 2013). An example of such mechanisms, also endorsed by S&T, are mandatory “cooling-off” periods that aim to induce subjects to critically reconsider their own decisions. This has been used, for instance, before sales of financial products

²⁵The judgements that people form can be either “discovered” (Plott 1996) or “constructed” (Slovic 1995). Ideal conditions should favour the formation of “considered” judgements (and other subjective attitudes) that are stable, at least temporarily, after a process of experience and reflection (see Gauthier 1986, ch. 2; cf. Sugden 2006, pp. 211-9).

are confirmed. Similar mechanisms can be used in the design of methods to elicit subjective attitudes for use in policy analysis (e.g. in health and environmental domains). For instance, the use of interactive designs in which individuals are asked to reflect upon their choices has the potential to induce individuals to form considered subjective attitudes that mitigate some of the problems caused by choice reversals (see also Bleichrodt et al. 2001, p. 1499 and Slovic 1995, pp. 369-70). Other examples include providing information to individuals before taking their choices into account (Elster 1982, p. 221), promoting further rights of education and non-manipulation (Arneson 1990, p. 168), requiring or encouraging active choices (Klevmarken 2002; Hedlin and Sunstein 2016), and providing educational opportunities to agents that regret their past behaviour (Gilboa 2010).

These policies are not desirable because they are assumed to benefit bounded rational agents while imposing minimal costs on “fully-rational” agents (Camerer et al. 2003, pp. 1211-2), but rather because they have the potential to benefit agents that do not endorse their behaviour while imposing minimal costs on other agents. These policies also differ from many policies proposed by S&T and Camerer et al. (2003) in that they share the feature of imposing very minimal restrictions upon the targeted agents’ *positive freedom* (see fn. 22).

6. Concluding Remarks

When data on choice behaviour is available and conflicting choice patterns are present, active observers are faced with the task of discriminating between choices for welfare or policy evaluation. According to the arguments presented here, an observer should use choices that individuals confirm for this end. I have also argued that confirmed choices have decisive advantages over context-independent and reason-based choices as proxies of welfare. Finally, I have made the case that even if one supports a fully-informed hypothetical notion of well-being, confirmations are still necessary for choices to be reliable proxies of welfare and guide policy.

This is surely at a cost. Asking people to evaluate their choices is an additional layer of data to be gathered. Sometimes this will be easily implemented, but some others it will be difficult or even impossible. The cost will be less compared to the benefits the more unconfirmed choices are frequent. Throughout the paper I have often used simple examples of unconfirmed choices as illustrations. However, it is not only in exceptional cases that people do not confirm their behaviour. Casual observation and introspection suggest that people often regret, change their minds, and/or have conflicting motivations concerning minor and major decisions in their lives. Benjamin et al. (2012) provide recent empirical data to back this observation. They asked subjects both what they think they would choose and what they would themselves want to choose in a series of hypothetical choice scenarios. For instance, one of their hypothetical choice scenarios was between a job in which the subject would “sleep more but earn less” and a job in which the subject would “sleep less but earn more”. For each scenario, they asked subjects the

two following questions: “If you were limited to these two options, which do you think you would choose?” (“stated choice”), followed by “If you were limited to these two options, which would you want yourself to choose?” (“stated meta-choice”). In a total of 7302 pairs of observations in a sample of 929 subjects, 28% of subjects’ stated choices differed from their stated meta-choices, providing evidence that many individuals do not confirm the choices they expect to make.

Given the potential contingency of choices, preferences, and judgements on changes in view-point or context, it will sometimes be difficult to have credible rankings based on subjective information. My arguments intend to support that we may not want to forget about this information altogether, but try instead to have richer and more reliable data sets that include information on both choices and confirmations. Without the presumption of being able to recover stable and context-independent latent preferences, we may still be able to recover choices and judgements that are meaningful for normative analysis for a given context and time.

References

- Adler, M. D. and P. Dolan (2008) Introducing a ‘Different Lives’ Approach to the Valuation of Health and Well-Being. University of Pennsylvania Law School Institute for Law and Economics Research Paper 08-05.
- Alós-Ferrer, C. and F. Strack (2014) From Dual Processes to Multiple Selves: Implications for Economic Behavior. *Journal of Economic Psychology* 41: 1–11.
- Apestequia, J. and M. A. Ballester (2015) A Measure of Rationality and Welfare. *Journal of Political Economy* 123(6): 1278–1310.
- Ariely, D., G. Loewenstein, and D. Prelec (2003) ‘Coherent Arbitrariness’: Stable Demand Curves Without Stable Preferences. *The Quarterly Journal of Economics* 118: 73–105.
- Arneson, R. J. (1990) Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare. *Philosophy and Public Affairs* 19(2): 158–94.
- Ashraf, N., D. Karlan, and W. Yin (2006) Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines. *The Quarterly Journal of Economics* 121(2): 635–72.
- Augenblick, N., M. Niederle, and C. Sprenger (2015) Working Over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics* 130(3): 1067–115.
- Baigent, N. (1995) Behind the Veil of Preferences. *Japanese Economic Review* 46(1): 88–101.
- Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2012) What Do You Think Would Make You Happier? What Do You Think You Would Choose?. *The American Economic Review* 102(5): 2083–110.
- Benjamin, D. J., O. Heffetz, M. S. Kimball, and N. Szembrot (2014) Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference. *The American Economic Review* 104(9): 2698–735.
- Berlin, I. (1969) *Four Essays on Liberty*. Oxford University Press., Oxford.
- Bernheim, B. D. (2009) Behavioral Welfare Economics. *Journal of the European Economic Association* 7(2-3): 267–319.

- Bernheim, B. D. and A. Rangel (2004) Addiction and Cue-Triggered Decision Processes. *The American Economic Review* 94(5): 1558–90.
- (2007) Toward Choice-Theoretic Foundations for Behavioral Welfare Economics. *The American Economic Review: Papers and Proceedings* 97(2): 464–70.
- (2009) Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *The Quarterly Journal of Economics* 124(1): 51–104.
- Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2008) How are Preferences Revealed?. *Journal of Public Economics* 92: 1787–94.
- Birnbaum, M. H. and U. Schmidt (2015) The Impact of Learning by Thought on Violations of Independence and Coalescing. *Decision Analysis* 12(3): 144–52.
- Bleichrodt, H., J. L. Pinto, and P. P. Wakker (2001) Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. *Management Science* 47(11): 1498–514.
- Bonein, A. and L. Denant-Boèmont (2015) Self-control, Commitment and Peer Pressure: A Laboratory Experiment. *Experimental Economics* 18: 543–68.
- Botti, S. and A. L. McGill (2006) When Choosing Is Not Deciding: The Effect of Perceived Responsibility on Satisfaction. *Journal of Consumer Research* 33(2): 211–19.
- Brandt, R. (1979) *A Theory of the Good and the Right*. Oxford University Press, Oxford.
- Bratman, M. E. (1996) Identification, Decision, and Treating as a Reason. *Philosophical Topics* 24(2): 1–18.
- (2003) A Desire of One’s Own. *The Journal of Philosophy* 100(5): 221–42.
- Broome, J. (1978) Choice and Value in Economics. *Oxford Economic Papers* 30(3): 313–33.
- Camerer, C. F., G. Loewenstein, and D. Prelec (2004) Neuroeconomics: Why Economics Needs Brains. *Scandinavian Journal of Economics* 106: 555–79.
- Camerer, C., S. Issacharoff, G. Loewenstein, T. O’Donoghue, and M. Rabin (2003) Regulation for Conservatives: Behavioral Economics and the Case for ‘Asymmetric Paternalism’. *University of Pennsylvania Law Review* 151: 1211–54.
- Center for Disease Control and Prevention (2011) 70 Percent of Smokers Want to Quit as Nation Approaches the Great American Smokeout. https://www.cdc.gov/media/releases/2011/p1110_smoking_treatment.html.
- Chambers, C. P. and T. Hayashi (2012) Choice and Individual Welfare. *Journal of Economic Theory* 147: 1818–49.
- Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2004) For Better or For Worse: Default Effects and 401 (k) Savings Behavior. In: D. A. Wise (ed) *Perspectives on the Economics of Aging*. University of Chicago Press, Chicago and London: 81–126.
- Cowen, T. (1993) The Scope and Limits of Preference Sovereignty. *Economics and Philosophy* 9: 253–69.
- Cubitt, R. P., C. Starmer, and R. Sugden (2001) Discovered Preferences and the Experimental Evidence of Violations of Expected Utility Theory. *Journal of Economic Methodology* 8(3): 385–414.
- Deaton, A. S., D. Kahneman, A. Krueger, D. Schkade, N. Schwarz, and Arthur Stone (2011) Memo to the Office of National Statistics’ Advisory Group on Subjective Well-Being. In: *Supporting Documents for the Meeting to Provide Guidance to the Organisation for Economic Cooperation and Development on its Plans to Measure Self-Reported Well-Being*. Organisation for Economic Co-operation and Development (OECD), Paris.

- Decancq, K., M. Fleurbaey, and E. Schokkaert (2015) Happiness, Equivalent Incomes and Respect for Individual Preferences. *Economica* 82: 1082–106.
- Elster, J. (1982) Sour Grapes — Utilitarianism and the Genesis of Wants. In: A. K. Sen and B. Williams (eds) *Utilitarianism and Beyond*. Cambridge University Press, Cambridge: 219–38.
- Fehr, E. and K. Hoff (2011) Introduction: Tastes, Castes and Culture: The Influence of Society on Preferences. *The Economic Journal* 211: 396–412.
- Frankfurt, H. G. (1971) Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1): 5–20.
- (1988) Identification and Wholeheartedness. In: *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.
- Fudenberg, D. and D. K. Levine (2006) A Dual Self Model of Impulse Control. *The American Economic Review* 96: 1449–76.
- (2012) Timing and Self-Control. *Econometrica* 80(1): 1–42.
- Fumagalli, R. (2013) The Futile Search for True Utility. *Economics and Philosophy* 29: 325–47.
- (2016) Decision Sciences and the New Case for Paternalism: Three Welfare-related Justificatory Challenges. *Social Choice and Welfare* 47: 459–80.
- Gauthier, D. (1986) *Morals by Agreement*. Clarendon Press, Oxford.
- George, D. (1984) Meta-Preferences: Reconsidering Contemporary Notions of Free Choice. *International Journal of Social Economics* 11(3/4): 92–107.
- Gilboa, I. (2010) Questions in Decision Theory. *Annual Review of Economics* 2(1): 1–19.
- Gilboa, I. and D. Schmeidler (2001) *A Theory of Case-Based Decisions*. Cambridge University Press, Cambridge, UK.
- Gul, F. and W. Pesendorfer (2008) The Case for Mindless Economics. In: A. Caplin and A. Schotter (eds) *The Foundations of Positive and Normative Economics*. Oxford University Press, New York: 3–39.
- Harsanyi, J. C. (1997) Utilities, Preferences, and Substantive Goods. *Social Choice and Welfare* 14: 129–45.
- Hausman, D. M. (2012) *Preference, Value, Choice, and Welfare*. Cambridge University Press, New York.
- (2013) A reply to Lehtinen, Teschl and Pattanaik. *Journal of Economic Methodology* 20(2): 219–23.
- (2016) On the Econ within. *Journal of Economic Methodology* 23(1): 26–32.
- Hausman, D. M. and M. S. McPherson (2009) Preference Satisfaction and Welfare Economics. *Economics and Philosophy* 25: 1–25.
- Hedlin, S. and C. R. Sunstein (2016) Does Active Choosing Promote Green Energy Use? Experimental Evidence. *Ecology Law Quarterly* 43(1): 107–42.
- Hirschman, A. O. (1984) Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse. *The American Economic Review: Papers and Proceedings* 74(2): 89–96.

- Hoff, K. and J. E. Stiglitz (2016) Striving for Balance in Economics: Towards a Theory of the Social Determination of Behavior. *Journal of Economic Behavior and Organization* 126: 25–57.
- Infante, G., G. Lecouteux, and R. Sugden (2016) Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics. *Journal of Economic Methodology* 23(1): 1–25.
- Jeffrey, R. C. (1974) Preferences Among Preferences. *The Journal of Philosophy* 71(13): 377–91.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. Farrar, Straus & Giroux, New York, NY.
- Kahneman, D. and A. S. Deaton (2010) High Income Improves Evaluation of Life but not Emotional Well-Being. *Proceedings of the National Academy of Sciences* 107(38): 16489–93.
- Kahneman, D. and J. Riis (2005) Living, and Thinking About It: Two Perspectives on Life. In: F. A. Huppert, B. Kaverne, and N. Baylis (eds) *The Science of Well-being*. Oxford University Press, : 285–304.
- Kahneman, D. and A. Tversky (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47(2): 263–92.
- Kahneman, D., P. Wakker, and R. Sarin (1997) Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics* 112: 375–406.
- Kirman, A. and M. Teschl (2006) Searching for Identity in the Capability Space. *Journal of Economic Methodology* 13(3): 299–325.
- Klevmarken, N. A. (2002) Swedish Pension Reforms in the 1990s. Working Paper, No. 2002:6, Uppsala University, Department of Economics.
- Koszegi, B. and M. Rabin (2007) Mistakes in Choice-based Welfare Analysis. *American Economic Review Papers and Proceedings* 97(2): 477–81.
- van de Kuilen, G. (2009) Subjective Probability Weighting and the Discovered Preference Hypothesis. *Theory and Decision* 67: 1–22.
- van de Kuilen, G. and P. P. Wakker (2006) Learning in the Allais Paradox. *Journal of Risk and Uncertainty* 33: 155–64.
- Lemaire, S. (2016) A Stringent but Critical Actualist Subjectivism about Well-Being.. *Les ateliers de l'éthique* 11(2): 113–50.
- Lewis, D. (1989) Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Supplementary Volumes* 63: 113–137.
- Lippert-Rasmussen, K. (2003) Identification and Responsibility. *Ethical Theory and Moral Practice* 6: 349–76.
- Little, I. M. D. (1949) A Reformulation of the Theory of Consumer's Behaviour. *Oxford Economic Papers* 1(1): 90–9.
- Manzini, P. and M. Mariotti (2014) Welfare Economics and Bounded Rationality: The Case for Model-based Approaches. *Journal of Economic Methodology* 21(4): 343–60.
- Mill, J. S. (1859) *On Liberty*. Penguin Classics, London[2010] edition.
- Nishimura, H. (2018) The Transitive Core: Inference of Welfare from Nontransitive Preference Relations. *Theoretical Economics* 13: 579–606.
- Noggle, R. (1999) Integrity, the Self, and Desire-Based Accounts of the Good. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 96(3): 303–31.

- Noor, J. (2011) Temptation and Revealed Preference. *Econometrica* 79: 601–44.
- Office for National Statistics (2017) Adult Smoking Habits in the UK 2017. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2017>.
- Olsaretti, S. (2006) Introduction. In: S. Olsaretti (ed) *Preferences and Well-being*. Cambridge University Press, Cambridge, UK: 1–7.
- Parfit, D. (1984) *Reasons and Persons*. Oxford University Press, Oxford.
- (2011) *On What Matters: Volume 1*. Oxford University Press, New York.
- Plott, C. R. (1996) Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis. In: K. Arrow, E. Colombatto, M. Perlman, and C. Schmidt (eds) in *The Rational Foundations of Economic Behavior*. Basingstoke: Macmillan and International Economic Association, : 225–50.
- Qizilbash, M. (2012) Informed Desire and the Ambitions of Libertarian Paternalism. *Social Choice and Welfare* 38: 647–58.
- Rabin, M. (2013) Incorporating Limited Rationality into Economics. *Journal of Economic Literature* 51(2): 528–43.
- Read, D. (2006) Which Side Are You On? The Ethics of Self-command. *Journal of Economic Psychology* 27: 681–93.
- Read, D. and B. van Leeuwen (1998) Predicting Hunger: The Effects of Appetite and Delay on Choice. *Organizational Behavior and Human Decision Processes* 76(2): 189–205.
- Rosati, C. (1995) Persons, Perspectives, and Full Information Accounts of the Good. *Ethics* 105: 296–325.
- Rubinstein, A. and Y. Salant (2012) Eliciting Welfare Preferences from Behavioural Data Sets. *The Review of Economic Studies* 79: 375–87.
- Ryff, C. D. (1989) Happiness is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being. *Journal of Personality and Social Psychology* 57(6): 1069–81.
- Sagoff, M. (1986) Values and Preferences. *Ethics* 96(2): 301–16.
- Salant, Y. and A. Rubinstein (2008) (A, f): Choice with Frames. *The Review of Economic Studies* 75(4): 1287–96.
- Samuelson, P. A. (1963) Discussion. *The American Economic Review: Papers and Proceedings* 53(2): 227–36.
- Scanlon, T. M (2002) Reasons and Passions. In: S. Buss and L. Overton (eds) *Contours of Agency: Essays on Themes from Harry Frankfurt*. MIT Press, Cambridge MA: 165–83.
- Sen, A. K. (1977) Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy and Public Affairs* 6(4): 317–44.
- Sidgwick, H. (1907) *The Methods of Ethics*. Hackett, Indianapolis[7th ed.] edition.
- Slovic, P. (1995) The Constitution of Preference. *American Psychologist* 50(5): 364–71.
- Sobel, D. (1994) Full Information Accounts of Well-Being. *Ethics* 104: 784–810.
- (2009) Subjectivism and Idealization. *Ethics* 119(2): 336–52.
- Starmer, C. and R. Sugden (1989) Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect. *Journal of Risk and Uncertainty* 2: 159–78.

- Sugden, R. (2004) The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences. *The American Economic Review* 94(4): 1014–33.
- (2006) Taking Unconsidered Preferences Seriously. In: S. Olsaretti (ed) *Preferences and Well-Being*. Cambridge University Press, Cambridge, UK: 209–32.
- (2008) Why Incoherent Preferences do not Justify Paternalism. *Constitutional Political Economy* 19(3): 226–48.
- (2018) *The Community of Advantage: A Behavioural Economist’s Defense of the Market*. Oxford University Press, Oxford.
- Sunstein, C. R. and R. H. Thaler (2003) Libertarian Paternalism Is Not an Oxymoron. *The University of Chicago Law Review* 70(4): 1159–202.
- Thaler, R. H. and H. M Shefrin (1981) An Economic Theory of Self Control. *Journal of Political Economy* 89(2): 392–406.
- Thaler, R. H. and C. R. Sunstein (2003) Libertarian Paternalism. *The American Economic Review* 93: 175–9.
- (2008) *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, New Haven, CT.
- Tversky, A. and D. Kahneman (1974) Judgment Under Uncertainty: Heuristics and Biases. *Science* 185: 1124–31.
- (1981) The Framing of Decisions and the Psychology of Choice. *Science* 211: 453–8.
- Velleman, J. D. (1999) A Right to Self-termination?. *Ethics* 109: 606–28.
- Voorhoeve, A. (2013) Response to Rabin. In: A. Oliver (ed) *Behavioural Public Policy*. Cambridge University Press, : 140–47.
- Waldfoegel, J. (2005) Does Consumer Irrationality Trump Consumer Sovereignty?. *Review of Economics and Statistics* 87(4): 691–96.
- Watson, G. (1975) Free Agency. *The Journal of Philosophy* 72: 205–20.
- World Health Organization (2018) Tobacco Use is One of the Chief Preventable Causes of Death in the World. <http://www.who.int/news-room/fact-sheets/detail/tobacco>.