Which choices merit deference? A comparison of three behavioural proxies of subjective welfare

João V. Ferreira*

Abstract

Recently several authors have proposed proxies of welfare that equate some (as opposed to all) choices with welfare. In this paper, I first distinguish between two prominent proxies: one based on *context-independent choices* and the other based on *reason-based choices*. I then propose an original proxy based on choices that individuals state they would want themselves to repeat at the time of the welfare/policy evaluation (*confirmed choices*). I articulate three complementary arguments that, I claim, support confirmed choices as a more reliable proxy of welfare than context-independent and reason-based choices. Finally, I discuss the implications of these arguments for *nudges* and *boosts*.

JEL classification: D01; D04; D60; D90; I31.

Keywords: Behavioural welfare economics; Confirmed choices; Stated meta-choices; Nudges; Boosts.

1 Introduction

It has been standard practice in neoclassical economics to assume a tight link between choice and individual (subjective) welfare. This link has usually been justified on three different grounds: (i) choices are said to reveal preferences that individuals want to be satisfied, (ii) choices are said to be a good proxy of subjective well-being, and/or (iii) choices are said to be worth respecting for the sake of individual sovereignty.[1] Following these justifications, individual choices are said to provide 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. However, evidence from psychology

*University of Southampton, SO17 1BJ Southampton, UK. Email: j.ferreira@soton.ac.uk URL: https://joaovferreira.weebly.com

[1] See, e.g., Little (1949), Samuelson (1963), and Gul and Pesendorfer (2008) for a mix of these claims (see Bernheim 2009: 290-3 for a review). For example, Little (1949: 98) invokes (ii) and (iii) when he argues 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". See, e.g., Sen (1973), Broome (1978), and Hausman and McPherson (2009) for critical reviews.

and behavioural economics on preference reversals, framing effects, and problems of self-control, among other behavioural phenomena, show 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" (e.g. Tversky and Kahneman 1974; Ariely et al. 2003; Choi et al. 2004; cf. Fudenberg et al. 2012; Maniadis et al. 2014): Even major decisions about retirement savings can be influenced by whether the option to enrol in a high-contribution pension plan is set as an option from which one can opt out or set as an option to which one needs to opt in. These kinds of findings question whether individuals' actual choices provide a reliable welfare ranking of alternatives. To illustrate this problem, consider the following example:

> ***The snack choice.*** Suppose that Norah is offered the choice between an *apple* and a *snickers* bar. In a first situation (period 1), she is asked to choose in advance which one she wants to consume one week later, and she chooses the *apple*. One week later (period 2), she is asked which one she wants to consume immediately, and she now chooses the *snickers*.[3]

Which of these choices, if either, provides a good indication of what is good for Norah as *actually judged by herself*? In this paper, I combine insights from behavioural economics, psychology, and moral philosophy to *compare three behavioural proxies of subjective welfare* that provide different insights into this kind of question. This paper is thus concerned, as the title indicates, with examining which choices merit deference (see Bernheim 2016: 43) or, in other words, with identifying a reliable "way to disrespect choice" (see Manzini and Mariotti 2014: 347).

First, I review two prominent behavioural proxies of welfare that equate some (as opposed to all) choices with welfare. The first of these argues that only choices that remain constant across context and time (*context-independent choices*) are reliable proxies of welfare. In the snack choice example, this proxy is agnostic about which snack is "better" for Norah. The second argues that only choices that are made after rational deliberation (*reason-based choices*) are reliable proxies of welfare. This approach demands that, alongside choice data, an observer gathers information about the origin of choices. For instance, in the snack choice example, suppose we have gathered evidence that immediate consumption triggers impulsiveness, while being far from the moment of consumption encourages slow and reasoned deliberation. Most versions of this proxy would infer that the apple is better for Norah as judged by herself.

[2] The literature on this subject is vast. See, e.g., Kahneman (2011), Rabin (2013), and Hoff and Stiglitz (2016) for reviews.

[3] This example is taken from the experiment by Read and van Leeuwen (1998), which relates to the prominent literature in economics on time-inconsistent behaviour between smaller short-term rewards and larger long-term rewards (see Rabin 2013 for a review). In the experiment, subjects are not aware that they can redo their choice a week later, and between 62% and 82% of subjects — depending on the treatments which differ in terms of current and future states of hunger — that chose a "healthy" snack in the advance choice situation reversed their choice for an "unhealthy" snack in the immediate choice situation.

As I will demonstrate below, however, these proxies are not robust to common phenomena such as updating beliefs, changing preferences, ex-post regret, taste for variety, habituation, and self-control issues like addiction. In response, I propose an original proxy that addresses these and other concerns, according to which only choices that individuals would want themselves to repeat at the time of the welfare or policy evaluation (*confirmed choices*) are reliable proxies of welfare. This approach demands that, alongside choice data, an observer records "stated meta-choices" at the time of the welfare/policy evaluation: that is, self-reports about what people would want themselves to choose at that point in time.[4] In the snack choice example, suppose that the observer's welfare evaluation takes place a week later at period 3 and that, for the sake of illustration, Norah reports at this period that she would want herself to choose the snickers over the apple if faced with either of the two choice situations at period 3. According to this proxy, the snickers should be deemed "better" for Norah at period 3 as judged by herself. In fact, her actual choice of the snickers instead of the apple seems to have revealed a ranking between the snickers and the apple that Norah (and not the observer) deems relevant for her welfare at the time of the welfare evaluation.

I provide three complementary arguments that support confirmed choices as a more reliable proxy of subjective welfare than context-independent and reason-based choices. First, I argue that confirmed choices are a more reliable proxy of preference satisfaction than context-independent and reason-based choices. In a nutshell, I distinguish between two types of preferences — ones that are, and ones that are *not* aligned with what is good for individuals as judged by themselves — and then argue that stated meta-choices are instrumental for identifying which choices reveal which type of preferences at the time of the welfare evaluation. Second, I argue that confirmed choices are a more reliable proxy of multidimensional *subjective well-being* (hereafter SWB) than context-independent and reason-based choices. This, I claim, follows from the previous argument as well as from the argument that stated meta-choices trace important aspects of SWB, such as living according to personal values and avoiding negative emotions like regret. Third, I argue that confirmed choices are more respectful of individual sovereignty than context-independent and reason-based choices. The main underlying idea is that the respect of individual sovereignty is *not* equivalent to the respect of all individual choices but instead, I claim, akin to the respect of only the choices that individuals want respected. Stated meta-choices at the time of the welfare or policy evaluation provide an indication of which choices individuals want respected at that particular point in time.

[4] I borrow the term "stated meta-choice" from Benjamin et al. (2012). Benjamin et al. (2012) faced 929 subjects with a series of hypothetical choice scenarios, such as a hypothetical choice 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, 28% of subjects' stated choices differed from their stated meta-choices. This evidence supports the prevalence of a conflict between what people (think they would) do and what they would want themselves to do.

These arguments have several policy implications. I explore their consequences for two influential behavioural policy programmes that emerged from research in psychology and behavioural economics: *nudges* and *boosts*.[5] The former are generally non-coercive and non-incentivised interventions that aim to steer people towards welfare-promoting behaviour by changing how choices are presented. I argue that stated meta-choices can be used to target nudges to individuals who do not confirm their behaviour, such as smokers who would like to quit smoking, and that such interventions have several advantages over traditional nudges that are not targeted to a sub-population. The latter, boosts, are generally educational interventions that foster people's decision-making competencies to help them reach their objectives. I argue that boosting individuals' ability and opportunities to reflect upon their past behaviour may help them to better reach their objectives, and I provide several examples of such interventions.

To avoid misunderstandings, it is worth emphasising that I do regard confirmed choices as a fallible proxy of subjective welfare. This is clear for long-term notions of welfare such as lifetime welfare, but it is also the case when restricting oneself to the period of the welfare/policy evaluation. For example, self-reports are sometimes vulnerable to deception, framing effects, and strategic concerns. Below, I will argue that — insofar as we are concerned with a proxy of *subjective welfare* — self-reports can reveal relevant information for observers' welfare/policy evaluations that is not captured in choice behaviour. However, it is important to recognize that these reports may provide unreliable information in the absence of favourable conditions for revealing self-reports that are honest, informed, reflected, and robust to trivial changes in viewpoint or context. This means that the way self-reports are revealed should be taken into consideration, and I discuss this issue when addressing potential objections below.

Before proceeding, it is also worth noting four simplifications of my inquiry. First, my inquiry is only concerned with *subjective welfare*, here understood as what is good for individuals as actually judged by themselves. This contrasts with objective welfare theories that are based on judgements about what is good for individuals that are either hypothetical (like in "informed preference" theories, as in, e.g., Brandt 1979 and Harsanyi 1997) or alien to the individual (like in "objective-list" theories, as in, e.g., Nussbaum 2006). This focus accords with the tradition in economics to treat individual subjective attitudes as decisive in assessing the relative welfare associated with different alternatives. Second, my inquiry is only concerned with *proxies* of subjective welfare, i.e., with observable data that can be used for an observer's welfare evaluation at a given point in time. My goal is thus related but independent from the goal of theories of welfare, which aim to identify which choices, preferences, or other constructs represent or ought to represent welfare (see, e.g., Olsaretti 2006). For instance, it is coherent to uphold a theory

[5]Numerous institutions around the world, inside and outside governments, implement policy interventions derived from these programmes (see, e.g., OECD 2017). See Grüne-Yanoff and Hertwig (2016) and Hertwig and Grüne-Yanoff (2017) for the features that distinguish these two behavioural policy programmes and their underlying research programmes.

of welfare according to which some hypothetical preferences ought to represent welfare, while at the same time holding that in general actual choices are the best proxy of welfare available (see, e.g., Arneson 1990: 164). Third, I only *compare* three proxies of welfare that use actual choices as their primary data. Thus, I do not directly address the question of which choices are the "best" proxy of welfare. By comparing three behavioural proxies, I aim to (i) expose the limitations of currently held proxies, (ii) propose a reliable (even if fallible) alternative, and (iii) enrich the debate on the criteria that might be used to identify which choices merit deference. Fourth, I do not address the question of whether choices are the "right" proxy of welfare. Many authors believe that they are not, and some have proposed alternative proxies of welfare based on individuals' experiences of pleasure and pain or on the activation patterns of specific areas in the brain (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 inferences is, however, a common practice in economics. There is therefore a pragmatic reason for identifying which choices are a reliable proxy of subjective welfare (see also Chambers and Hayashi 2012 and Manzini and Mariotti 2014).

Finally, it is worth mentioning some concepts that, although not always applied to welfare analysis, seem related to confirmed choices and stated meta-choices. One example is what philosophers often call *second-order desires or volitions* (e.g. Frankfurt 1971; Jeffrey 1974). For instance, "I want not to want to smoke" is a second-order desire/volition. Stated meta-choices can be seen as an observable measure of second-order desires/volitions. I diverge from previous authors by focusing on this observable measure and using it as an input to a proxy of subjective welfare. Stated meta-choices are also related to but different from the concept of *meta-preferences* (e.g. Sen 1977; George 1984). Meta-preferences are usually assumed to be a single and stable ordering of multiple preferences defined over the universe of alternatives. In my analysis choices are the primitive data (as opposed to multiple preferences), and stated meta-choices are *not* assumed to be stable over time. My analysis is also fully devoted to (behavioural) welfare analysis, while the previous analyses are not. Likewise, confirmed choices are related to but different from the standard interpretation of *meta-choices* (e.g. Bernheim and Rangel 2009: 83). While a choice is said to be confirmed as long as the individual self-reports that she would want herself to repeat it at the time of the welfare or policy evaluation, a meta-choice is usually assumed to be a choice made in advance between two or more choices. Lastly, the confirmed choice notion resembles rational choice notion proposed by Gilboa and Schmeidler (2001: 17-8): "An action, or a sequence of actions is rational for a decision maker if, 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). Besides the noticeable difference of domain (welfare versus rationality), a confirmed choice is said to be confirmed based on a self-report that may or may not be informed by an *external analysis of the decisions involved* and/or *additional*

information. In addition, an individual may decide not to confirm a choice because of reasons other than regret.

The remainder of the paper is organised as follows. In Section 2, I describe the proxy based on context-independent choices. In so doing, I introduce the general framework for describing choice behaviour proposed by Bernheim and Rangel (2009) that I will be using throughout the paper to illustrate the welfare inferences of the different proxies. In Section 3, I present the proxy based on reason-based choices. After that, I develop the conceptual apparatus of the proxy based on confirmed choices (Section 4). In Section 5, I articulate three complementary arguments that support confirmed choices as a more reliable proxy of welfare than context-independent and reason-based choices. I then discuss some potential objections and provide insights on how to address them (Section 6). In Section 7, I discuss the implications of these arguments for the behavioural policy programmes that use nudges and boosts. Section 8 concludes.

2 Context-independent Proxy

There are several proposals for how to identify which choices merit deference that aim to reconcile welfare economics with recent behavioural findings. One of the proposals that comes 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 position, which I call the *context-independent proxy* of welfare, has been formalised by Bernheim and Rangel (2009) (hereafter B&R).[6] 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. To model context-dependent behaviour, B&R define a *generalised choice situation*, denoted $GCS = (A, d)$, as the combination of a standard *choice situation* $A \subseteq X$ and an *ancillary condition* d. An ancillary condition can be the way in which information is presented, the labelling of a particular option as the *status-quo*, or any other feature of the choice environment as long as it "may affect behaviour, but is not taken as relevant to a social planner's evaluation" (B&R: 55).

Let $\mathcal{G}^*$ denote the set of all GCSs contemplated by an *observer*[7], and assume that for all choice situations $A \in X$ there is some ancillary condition d such that $(A, d) \in \mathcal{G}^*$. Individual

[6] See 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, and Burghart et al. (2007) and Chetty et al. (2009) for empirical applications of B&R's framework. See Manzini and Mariotti (2014) for a criticism of B&R based on the fact that B&R's approach does not rely on an explicit model of decision making. See Chambers and Hayashi (2012) and Nishimura (2018) for further "model-less approaches" that rely upon choice behaviour to make welfare comparisons.

[7] B&R take the position of a social planner. Hereafter, an observer can also be an expert wishing to advise an individual or a mediator who seeks to facilitate a contract between parties. Observers are assumed to be impartial and benevolent.

behaviour is modelled through a choice correspondence $C : \mathcal{G}^* \Rightarrow X$, that assigns a non-empty set of alternatives $C(A, d) \subseteq A$ to every generalised choice situation $(A, d) \in \mathcal{G}^*$. An alternative $x \in C(A, d)$ is interpreted as an option that the individual *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 of $\mathcal{G}^*$, 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.

At this point, it is worth noting that B&R are agnostic about the process that gives rise to choices. In contrast 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: 52), or in other words, respecting choices is required to satisfy individual "self-determination" (Bernheim 2009: 290-3). *Prima facie*, their proxy seems like a natural extension of the principle of individual sovereignty to settings in which context-dependent behaviour is prevalent (see, however, Section 5.3 below).

This approach is appealing because it relies exclusively on choice data. However, 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 (Rubinstein and Salant 2012). In such circumstances, many pairs of alternatives x and y are not comparable under P^* (hereafter denoted xN^*y). In answer to this criticism, B&R deviate from their context-independent proxy and propose to "prune" $\mathcal{G}^*$ by using non-choice data to delete "suspect" GCSs. This refinement allows them to identify a *welfare-relevant domain*, $\mathcal{G} \subseteq \mathcal{G}^*$, consisting of all GCSs that merit deference and from which the observer takes normative guidance. Note that when referring to the context-independent proxy, I mean the criterion presented above that does not take this refinement into account. In other words, the context-independent proxy does *not* prune $\mathcal{G}^*$ (i.e., $\mathcal{G} = \mathcal{G}^*$). On the other hand, the next two proxies I will discuss can be represented as criteria for pruning $\mathcal{G}^*$, and I will illustrate their welfare inferences using this approach.[8]

[8] B&R's preferred method of pruning $\mathcal{G}^*$ relies on evidence gathered from psychology, neuroscience, and neuroeconomics on informational processing failures such as the incorrect use of information, lack of attention, or naive forecasting (pp. 83-5). Their refined proxy is in fact similar to the reason-based proxy revised in the next section.

Using B&R's framework and their context-independent proxy of welfare, the introductory example can be represented as follows:

Table 1: The snack choice (context-independent proxy)

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

The apple and the snickers are not comparable according to the context-independent proxy 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 proxy is agnostic about which snack is welfare superior because there are conflicting choice patterns.

3 Reason-Based Proxy

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, which 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.[9]

A prominent example of this *reason-based proxy* of welfare is given by "dual-system" models recently popularised by Kahneman (2011) (see Alós-Ferrer and Strack 2014 for a short review of different models). 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 the 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

[9] What exactly counts as "rational deliberation" is contestable and differs according to the approach/model. In this section, I present two different but for the most part compatible conceptions of rational deliberation derived from psychology and philosophy that are influential in economics. See Infante et al. (2016) for a critical review of this general approach.

are "mistakes". Choices made after "cold" deliberation — 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 in which Hausman (2012) aims to describe how the concepts of *preference, value, choice, and welfare* are and ought to be used in economics. In the book, Hausman argues that a preference is and ought to be an *evaluation* in the sense that it is the result of a rational deliberation about what one has most reason to do. This means that, according to Hausman's view, a person's choice that is not based on rational deliberation about what she has most reason to do does not reveal a preference. It follows that such a choice cannot be used as a proxy of welfare. Only reason-based choices are *potential* proxies of welfare.[10]

B&R's framework can be used to illustrate the welfare inferences of the reason-based proxy of welfare. Using this approach, data on internal deliberation prior to choosing can be used to delete certain GCSs. For example, if internal deliberation prior to choosing is "too fast", then the corresponding GCS is excluded. In practice, it is often difficult to determine whether or not internal deliberation is rational in specific choices made by specific individuals, so observers rely on indirect data gathered from studies in psychology, neuroscience, and other fields (e.g., eye-tracking or neuroimaging studies).

With these premises in mind, we can revisit the snack choice example. Evidence suggests that, on the one hand, immediate food consumption generally triggers impulsive behaviour (or System 1), while, on the other hand, being far from consumption generally encourages reason-based deliberation (or System 2) (e.g. Read and van Leeuwen 1998). According to the reason-based proxy, this information suggests that the immediate choice should be excluded from the welfare-relevant domain $\mathcal{G}$ and that the apple is welfare superior to the snickers as judged by Norah. This is represented in Table 2.

Table 2: The snack choice (reason-based proxy)

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

[10]Hausman (2012) argues that preferences are *total subjective comparative evaluations* which are only reliable proxies of welfare when they are self-interested, informed, and competently considered (see also Hausman and McPherson 2009; Hausman 2016). This view is related but significantly different from the informed preference theories that define a person's well-being as the satisfaction of the desires that the person would have if she had all relevant information and made full rational use of this information (e.g. Brandt 1979; Arneson 1990; Harsanyi 1997). For my analysis, the relevant distinction is that the reason-based approach, including that of Hausman (2012), aims to identify which *actual* choices merit deference, while informed preference theories aim to identify what individuals would choose in hypothetical, idealised situations. See, e.g., Cowen (1993), Sobel (1994, 2009), Rosati (1995), and Noggle (1999) for critical reviews of informed preference theories of welfare.

4 Confirmed Proxy

I have distinguished two prominent behavioural proxies of welfare that equate some (as opposed to all) choices with welfare. In this section, I propose a new proxy of subjective welfare: the *confirmed proxy* of welfare.

Consider a discrete time horizon $\mathcal{T} = \{1, ..., T\}$ and assume, without loss of generality, that all generalised choice situations $(A, d) \in \mathcal{G}^*$ are ordered in time from 1 to $T-1$. Assume that the observer makes his/her welfare or policy evaluation at period T. Assume as well that the observer wants the welfare or policy evaluation to be reliable at period T.[11] Then, for any period $t \in \mathcal{T}$:

Definition 1 (Confirmed choice). *An individual is said to confirm at T her choice of $x \in C(A, d)$ made at $t < T$ if and only if at T she would want herself to select x if faced with $C(A, d)$ at T.*

In other words, a choice is said to be confirmed whenever an individual would want herself to repeat that choice at the time of the welfare or policy evaluation if faced with the same menu and the same ancillary conditions. In practice, this proxy demands that, alongside choice data, an observer records self-reports at T about what an individual would want herself to do at T. I call such self-reports *stated meta-choices*. This can be elicited using questions at period T, such as: "From where you stand now, would you want yourself to choose the same alternative again?" Stated meta-choices should not be confused with stated choices, which correspond to what an individual thinks she would choose. This difference is relevant, as many people, in many circumstances, would like to behave differently from how they think they would behave (as, e.g., in cases of limited self-control; see Benjamin et al. 2012 for empirical evidence).

We can use B&R's framework to represent some welfare inferences of this proxy. In this case, "who" prunes $\mathcal{G}^*$ is no longer the observer but the individual herself. Take again the snack choice example. Recall that at period 3 (after the two choices have been made) Norah would want herself to repeat her choice of the snickers over the apple but not her choice of the apple over the snickers. Table 3 represents the snack choice example according to the confirmed proxy of welfare.

[11] As argued below, this assumption is desirable for the observer to make reliable welfare inferences that respect individual attitudes at the time of the welfare/policy evaluation. In some situations, however, this will not be possible, and the observer wants the welfare or policy evaluation to be reliable before or after T. I address these issues in Section 6. Note that a similar analysis can be made if the evaluation occurs at a period 0, where choices are predicted rather than observed. See Cerigioni (2017, 2020) and Ferreira and Gravel (2020) for frameworks related to B&R that explicitly introduce a chronological order of subsets.

Table 3: The snack choice (confirmed proxy)

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

In this example, the confirmed proxy provides an unambiguous welfare inference in favour of snickers over apple for period T.[12] However, this is not always the case. Sometimes, the confirmed proxy provides a *prudent* inference. To see this, consider the case of addictive behaviour represented in the following payoff table taken from Dalton and Ghosal (2012: 594):

Table 4: The smoking choice

	h_1	h_2
a_1	1	-1
a_2	2	0

where a_2 corresponds to *smoking* and a_1 corresponds to *not smoking*, and h_i represents Norah's health states (h_1 being healthier than h_2). Assume that Norah chooses repeatedly from this set of feasible actions, a_1 and a_2, in a "long-run" horizon from periods 1 up to $T-1$ (see Dalton and Ghosal 2012: 588-93). Suppose as well that she believes, wrongly, that her health state is stable over time. It follows that she always prefers to smoke at the beginning of the time horizon (a_2 is the dominant action for each h). However, in the long-run Norah's health state deteriorates to h_2 and the unique long-run outcome is (a_2, h_2) with a payoff of 0.[13] Now suppose that in period T Norah states that she would like to quit smoking.[14] The confirmed proxy uses this piece of information to make more reliable comparisons of subjective welfare at period T. This self-report reveals that Norah's observed behaviour, even though it is context-independent, is not aligned with how she would want herself to behave at T. According to the

[12] The snack choice example shows that pruning $\mathcal{G}^*$ with stated meta-choices at T can lead to a welfare ranking that is more discerning than one based on context-independent choices. It is worth noting, however, that P^* is not necessarily acyclic when using stated meta-choices to prune $\mathcal{G}^*$. This may be problematic because it may not be possible to identify welfare optima for some choice situations. Yet, this limitation is shared by most refinements of B&R's framework. In fact, even B&R's preferred welfare ranking P^* is only acyclic when one observes the individual choosing from at least all two-element and three-element subsets of X, which in practice rarely occurs. B&R's framework is used here to illustrate, in an accessible and commensurable way, possible inferences of alternative proxies of welfare. It would be possible, however, to use stated meta-choices at T as auxiliary data in other frameworks. For example, Chambers and Hayashi (2012) propose a mapping from stochastic choice to a transitive and complete welfare binary relation based on a few axioms and weights on every (*chosen alternative, choice situation*) pair. It would be possible to use stated meta-choices to select "reasonable" weights for the different (*chosen alternative, choice situation*) pairs.

[13] If Norah takes the feedback from actions to health states into account, then she always chooses a_1 and the unique long-run outcome is (a_1, h_1) with a payoff of 1 (Dalton and Ghosal 2012: 588-93).

[14] Addiction, like smoking, is a typical example of a conflict between what people do and what they would want themselves to do that involves high stakes. According to a recent report from the U.K.'s Office for National Statistics (2019), in Great Britain more than half (52.7%) of people aged 16 years and above who currently smoke said they wanted to quit.

confirmed proxy, given that there is a conflict between choices and stated meta-choices at T, an observer should take an agnostic position at T as to whether smoking is better for Norah than not smoking *as actually judged by herself.*

It is worth noting, however, that the confirmed proxy can inform policy interventions even when it provides a prudent inference. For example, assume that a social planner is considering policies on smoking. The planner would like to set a policy framework that takes into account people's choices and the inference that many smokers, like Norah, would like to quit smoking. The planner should not prohibit smoking, as this would be against smokers' revealed preferences. At the same time, the planner should not only not prohibit smoking *but also* provide opportunities to "unwilling" smokers to follow the behaviour that they would want themselves to follow (i.e., *not smoking*). Providing free consultations with specialised doctors is a policy in that direction. Since the two policies — non-prohibition and free consultations — are not mutually exclusive, the planner would respect choices *and* stated meta-choices and potentially improve the welfare of smokers that would like to quit smoking.

5 Why Confirmed Choices

In my opening remarks, I mentioned three justifications that have been traditionally used to support the link between choice and welfare. In this section, I revisit these justifications to support the link between confirmed choices and welfare. The overall argument can be summarised as follows:

> [1] If A is a more reliable proxy of preference satisfaction than B at t, if A is a more reliable proxy of SWB than B at t, and if A is more respectful of individual sovereignty than B at t, then A is a more reliable proxy of subjective welfare than B at t as far as the most common theories of subjective welfare in economics are concerned.
>
> [2] Confirmed choices are a more reliable proxy of preference satisfaction than context-independent and reason-based choices at T.
>
> [3] Confirmed choices are a more reliable proxy of SWB than context-independent and reason-based choices at T.
>
> [4] Confirmed choices are more respectful of individual sovereignty than context-independent and reason-based choices at T.
>
> – – – – – – – – – – –
>
> Therefore, confirmed choices are a more reliable proxy of subjective welfare than context-independent and reason-based choices at T as far as the most common theories of subjective welfare in economics are concerned.

In the following, I provide support for premises [2], [3], and [4] (Sections 5.1, 5.2, and 5.3 respectively). While doing so, I link my arguments to common phenomena of interest to economics, such as updating beliefs, changing preferences, ex-post regret, taste for variety, habituation, and limited self-control.

5.1 The Argument from Preference Satisfaction

In neoclassical economics, preference satisfaction is one of the main views of individual welfare.[15] 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 instead 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 potentially only some of those choices reveal preferences that are aligned with what is good for the individuals as judged by themselves (*welfare-enhancing preferences*). This accords with the view, held by many authors, that "revealed preferences often differ from normative preferences" (Beshears et al. 2008: 1787).

Since we are concerned with subjective welfare, the potential discrepancy between revealed and welfare-enhancing preferences should be judged by the individuals themselves. It follows that for an observer to identify welfare-enhancing preferences at T, it is essential to make a distinction between the preferences that individuals consider aligned with what is good for themselves at T and the preferences that individuals do not consider aligned with what is good for themselves at T. As pointed out by Sagoff (1986: 303), it is false that each person wishes her 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." Then, it seems that if an individual does not want a preference that she has revealed in the past to be satisfied at T, it follows that this past preference should not count as welfare-enhancing at T.

I argue that confirmed choices are a more reliable proxy of welfare-enhancing preference satisfaction at T than context-independent and reason-based choices. To see this, note that the context-independent and reason-based proxies are likely to keep choices within the welfare-relevant domain that reveal preferences that individuals do *not* consider aligned with what is good for themselves at T. Having made context-independent choices or having deliberated about reasons to act before choosing does not exclude the possibility of not wanting one's revealed preferences to be satisfied at a different point in time. For example, suppose that Norah faces a single-shot decision at period 1 between purchasing standard or premium travel insurance, and

[15] The claim that a person's well-being consists of the satisfaction of her desires (or some subset of them) is also influential in philosophy. Even 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).

assume that she chooses the standard insurance after a slow considered deliberation. Confronted with the unexpected responsibility she felt for the choice, at period 2 Norah wishes she had chosen premium instead.[16] Norah's choice, even though it was made after rational deliberation and it was not contradicted by another choice, has revealed a preference that, at period 2, is not welfare-enhancing as judged by herself. However, the context-independent and reason-based proxies would deem standard insurance better than premium insurance for Norah at period 2.

Conversely, the context-independent and reason-based proxies are likely to disregard choices that reveal preferences that individuals consider aligned with what is good for themselves at T. 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 1978: 326). It is not clear, then, why it is reasonable to disregard the preference-rankings revealed by these choices, as the reason-based proxy would do. In this case, the reason-based proxy is overly restrictive as it is likely to unduly reduce the number of choices in the welfare-relevant domain. A similar limitation holds for context-independent choices. A notable example is changes of mind (either belief updating or preference change). For instance, suppose that Norah used to choose *pork chops* over *veggie roast* in her local restaurant, but since 2021 she chooses *veggie roast* instead because she has formed a new ideal in favour of no animal suffering. According to the context-independent proxy, *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 deemed to be valuable or important (see Parfit 1984: ch. 8 for a related argument). Then, it seems that an observer should be able to infer that *veggie roast* is "better" for Norah since 2021.

The confirmed proxy can successfully accommodate all these cases. When a person considers that a revealed preference is not aligned with what is good for herself, the confirmed proxy recognizes, contrary to the other proxies, that it is ambiguous from an observer's point of view if the revealed preference is welfare-enhancing. When a person considers that a revealed preference is aligned with what is good for herself, this choice seems to deserve deference even if it is not context-independent and/or reason-based (at least from the point of view of subjective welfare, which is our focus here). The confirmed proxy follows this insight and includes this choice in the welfare-relevant domain.

These arguments support the claim that confirmed choices are a more reliable proxy of preference satisfaction at T than context-independent and reason-based choices (premise [2] above). It recognises the important distinction between revealed and welfare-enhancing preferences, and it is robust to common phenomena such as updating beliefs, changing preferences, and habituation that the other two proxies fail to accommodate.

[16] See Botti and McGill 2006 for the effect of perceived responsibility on post-choice satisfaction.

5.2 The Argument from Subjective Well-being

There has been a renewed interest in the notion and measurement of SWB (e.g. Kahneman et al. 1997; Kahneman and Riis 2005; Deaton et al. 2011; Benjamin et al. 2014). While the term SWB is often associated with happiness or life satisfaction, there is a growing recognition that SWB is multidimensional and governed by several "fundamental aspects" besides happiness or life satisfaction (e.g. Adler and Dolan 2008; Kahneman and Deaton 2010; Benjamin et al. 2014).

In this section, I argue that confirmed choices are a more reliable 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 Benjamin et al. 2012 for evidence supporting this claim). This seems especially to be the case when preference satisfaction is restricted to welfare-enhancing preferences. Then, a corollary to the previous argument is that confirmed choices are likely to be a more reliable proxy of SWB than context-independent and reason-based choices as far as preference satisfaction is concerned. I complement this argument by showing how stated meta-choices at T can trace other fundamental aspects of SWB such as living according to personal values, being *who* one wants to be, limited self-control, and negative emotions like regret.

Consider 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 a person's interest in having a meaningful, valuable, worthwhile life (see, 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). Stated meta-choices bring information, not captured by the context-independent and reason-based proxies, about people's values and goals, which are likely to be aligned with these aspects of SWB. For instance, people's values change over time — as in the example in Section 5.1 in which Norah chooses between *veggie roast* and *pork chops* — and stated meta-choices allow an observer to make an inference that is aligned with individuals' values at the time of the welfare or policy evaluation.

In terms of self-control, there is by now considerable evidence that some individuals are willing to self-impose commitments to limit self-control costs (e.g. Ashraf et al. 2006; Augenblick et al. 2015; Bonein and Denant-Boèmont 2015). For example, in a recent experiment that tested dynamically inconsistent preferences in effort, 59% of subjects committed to their initial effort allocation choice at price \$0 (Augenblick et al. 2015). On the one hand, it is reasonable to

assume that some individuals who are aware of their self-control problems will *not* confirm their choices of "tempting" options (e.g., a smoker who unwillingly relapses into smoking is not likely to confirm this choice). On the other hand, some individuals may well confirm their choices of "tempting" options (e.g., a student that is content with her procrastination habits is likely to confirm her choice to watch TV and delay studying for an exam). This means that the distinction between choices that are confirmed and choices that are not confirmed is relevant if an observer wants to make reliable welfare comparisons when tempting options are available. This approach seems more sensible, at least from a subjective welfare perspective, than to assume that all choices that are due to limited self-control are mistakes that should not be part of the welfare-relevant domain (as, e.g., in Bernheim and Rangel 2004).

Turning to negative emotions, avoiding them is often seen as an important aspect of SWB (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 ex-post regret ("I wish I had not done that") and melancholy ("I wish I had done that") can be revealed through stated meta-choices at T. For example, a consumer is very likely not to want herself to repeat at T the choice of buying a product that she regrets having bought, even if she had bought it after a slow and reasoned deliberation. A confirmed proxy, contrary to the context-independent and reason-based proxies, would exclude this choice from the welfare-relevant domain at T. This seems to accord with the consumer's well-being as judged by herself at T.

At the same time, confirmed choices do not trace important aspects of SWB such as positive emotions or aspirations that are unrelated to choice behaviour. A proxy based on confirmed choices also fails to provide any information on the quality and intensity of individuals' experiences and their memories of these (see Kahneman and Riis 2005). Note, however, that these concerns are shared by the proxies based on context-independent and reason-based choices.

In sum, although choices and stated meta-choices at T are certainly not sufficient for measuring SWB, taken together they trace aspects of well-being, such as personal values, goals, ex-post regret, and conflicts between what people do and what they would like themselves to do, that seem to escape proxies that rely on context-independent and reason-based choices. These advantages, as well as the argument from preference satisfaction (Section 5.1), provide support for premise [3] above.

5.3 The Argument from Individual Sovereignty

My last argument concerns the respect of individual sovereignty. In economics, individual sovereignty, often referred to as consumer sovereignty, has traditionally been seen as a grounding principle: individuals' subjective attitudes are treated as decisive in assessing the relative welfare associated with different states of affairs. In most cases, consumer sovereignty is associated

with the respect of individual choices. For example, Sugden (2004, 2018) argues for a concept of consumer sovereignty that attaches value to a person's opportunities to *act* as she wants.

Individual sovereignty is important regarding subjective welfare for at least two reasons. First, in many circumstances, individuals are better placed than third parties to identify which choices enhance their own well-being (see, e.g., Mill 1859 [2010]: ch. 4). For example, Waldfogel (2005) provides some evidence that this is the case for choices among consumption goods. Using data from the Christmas and Hanukkah gift-giving seasons, the author finds that individuals value their own purchases at an average of 18% more, per dollar spent, than they value gifts from their friends and family (excluding sentimental value). The respect of individual sovereignty, in this case, is regarded as instrumental to enhancing individuals' subjective welfare. Second, the reasons for deferring to a person's judgement go beyond her reliability as a judge (see, e.g., Mill 1859 [2010]: ch. 4; Velleman 1999: 608; Bernheim 2009: 291-3). Being capable of *self-determination* (i.e., being able to freely choose one's acts without external compulsion) is likely to be highly valued by many individuals. The respect of individual sovereignty, in this case, has a more direct (or intrinsic) value for subjective welfare.

I argue that confirmed choices are more respectful of individual sovereignty at T than context-independent and reason-based choices. The underlying idea is that the respect of individual sovereignty at T is *not* tantamount to the respect of individual choices. Instead, the respect of individual sovereignty at T is akin to the respect of choices that individuals want to be respected at T.[17] Stated meta-choices at T indicate which choices individuals want to be respected at that particular point in time.

The context-independent and reason-based proxies are, by contrast, likely to violate individual sovereignty. For the former, this claim may be surprising since B&R's main justification for respecting context-independent choices is the respect of individual self-determination (see also Bernheim 2009: 291-3). However, the respect of individual self-determination seems *not* to provide a rationale to rely exclusively on choices nor a rationale for why to select context-independent choices rather than others. For example, a person who regrets her past choices made at $t < T$ is likely not to want those choices respected at T, even if they are context-independent. A similar argument holds when conflicting choice patterns exist: An individual may want a choice to be respected at T even though it has been "contradicted" by another choice at some period $t < T$.

For reason-based choices, there seems to be no reason to suppose that only choices followed by rational deliberation respect individual sovereignty at T. Using a previous example, individuals may well want choices made out of habit to be respected, though they are not deliberated. Another example is an individual's taste for variety, which behaviourally translates into a conflicting

[17] I retain choices as the main ingredient of this revised principle of individual sovereignty given my focus on proxies of welfare that use actual choices as their primary data. One can formulate related principles based on subjective attitudes that are not necessarily revealed in choice behaviour (e.g. Decancq et al. 2015: 1083).

choice pattern. To respect individual sovereignty at T, the observer should be able to infer if the individual wants this conflicting choice pattern to be respected at T. Stated meta-choices at T provide a proxy for this judgement, while an external inference about rational deliberation seems unrelated to it.

These arguments suggest that in the presence of common phenomena such as habituation, ex-post regret, changes of mind, taste for variety, or other reasons for changes in subjective attitudes over time, confirmed choices will be more respectful of individual sovereignty at T than context-independent and reason-based choices, as stated in premise [4] above.

6 Discussion

The previous arguments suggest that confirmed choices have important advantages over context-independent and reason-based choices as proxies of welfare. These advantages are particularly prominent when assessing what is good for a person *as judged by herself at the time of the welfare or policy evaluation* (period T). The focus on a person's welfare as judged by herself accords with the tradition in economics of treating individual subjective attitudes as decisive in assessing the relative welfare associated with different alternatives. I depart from neoclassical economics, however, by requiring synchronicity between the observer's and the observed person's judgements. This requirement seems sensible as soon as we acknowledge, as often done in psychology, philosophy, and behavioural economics, that individuals' subjective attitudes change over time. Synchronicity is thus a necessary condition for respecting individual attitudes at the time of the welfare/policy evaluation, which I have argued to be an essential quality of a proxy of subjective welfare (Section 5.3). Requiring synchronicity is also important for the proxy of welfare to be robust to common behavioural phenomena such as updating beliefs, changing preferences, and ex-post regret. While the timing of an observer's evaluation is often overlooked in the literature, some authors have articulated views that point towards synchronicity between the observer's and the observed person's judgements. For example, Gul and Pesendorfer (2005: 433) argue that to determine whether a policy "improves the welfare of the agent it suffices to determine whether the agent would vote for the policy in the period in which it is introduced".

This raises the question, however, of whether confirmed choices at T are a reliable proxy of welfare for periods other than T.[18] To see this, consider the following example. Suppose that last Wednesday, at $T-1$, Norah went to a theatre matinee and loved it. This Wednesday, at T, Maria (a friend of Norah who is deciding what they will do that day) asks Norah if she would want herself to repeat the choice that day; Norah says no. However, the reason behind Norah's answer is that on the previous night she went binge-drinking with other friends and that today she does not feel like going to the theatre. Without knowing this last piece of information, Maria

[18]The discussion that follows has particularly benefited from the comments and suggestions of two anonymous referees.

takes Norah's answer into consideration and decides not to invite her to the theatre matinee that day. This seems to be a correct inference by Maria, the "planner", about what is best for Norah as judged by herself for that day (period T). However, Maria's inference may be incorrect for other periods. Next Wednesday, at $T+1$, Maria may decide not to invite Norah to the theatre when actually Norah would love to go.

This example illustrates the problem that may arise in instances of *intertemporal substitutability or complementarity* of a good with itself or other goods. Although this problem is not exclusive to the confirmed proxy, it is useful to understand how it impacts the reliability of its welfare inferences. On the one hand, as the example illustrates, this issue is not problematic for the confirmed proxy's welfare inferences if a policy is introduced at T and its consequences are restricted to T. On the other hand, this issue can be problematic if intertemporal substitutability/complementarity impacts stated meta-choice at T in ways that it does not affect other periods, as this makes the welfare inference especially contingent on T.

A similar issue may arise with *intertemporal choices*, i.e., "decisions in which the timing of costs and benefits are spread out over time" (Loewenstein and Thaler 1989: 181). Examples include how much schooling to obtain, how much to save for retirement, buying a house or choosing an insurance policy. While the confirmed proxy accounts for intertemporal implications before T that the other proxies fail to account for, the confirmed proxy may not account for intertemporal consequences after T. This will be the case if the stated meta-choices at T disregard these consequences or incorrectly account for future tastes (for theory and evidence on the incorrect account of future tastes, or "projection bias", see Loewenstein et al. 2003 and Frederick et al. 2002: 373).

Another potential objection to the confirmed proxy is that it gives a prominent (even if auxiliary) role to self-reports. Economists have traditionally been suspicious of self-reports. The usual criticism is that talk is cheap (e.g. Grüne-Yanoff 2012: 643). I agree. However, self-reports seem to reveal information about people's goals and values that is not captured in choice behaviour and that it is important for an observer's welfare or policy evaluation (see also Hirschman 1984; Lewis 1989; Beshears et al. 2008). For example, stated meta-choices that are contrary to choice behaviour seem to decrease our confidence that revealed preferences are welfare-enhancing; if a consumer writes a bad review of a product, we should be less confident that the product the consumer bought is good for her as judged by herself. In the confirmed proxy, self-reports help the observer to make an educated guess about individuals' attitudes towards their behaviour at the time of the welfare or policy evaluation (see Manzini and Mariotti 2014: 344 for an argument in favour of using non-choice data for a similar purpose).

Even so, self-reports are more reliable in some contexts than in others. In some typical economic settings, such as consumption and labour market behaviour, people are usually familiar with the context and they have few reasons to be dishonest or strategic in their answers about

what they would want themselves to do. However, this is not necessarily the case in other contexts. For instance, self-reports are likely to be unreliable when concerning criminal behaviour, tax-avoidance, or other forms of antisocial behaviour, since deception and/or self-deception are likely. In other contexts, such as difficult or unfamiliar decisions (e.g., some healthcare choices), self-reports, like choices, are likely to be unreliable due to false or incomplete information. In still further contexts, where self-reports may have a political impact, individuals may respond strategically. Perhaps even more generally, self-reports, like choices, may be susceptible to changes in viewpoint or framing.

The concerns highlighted in this section suggest that in some contexts it may be important to create *favourable conditions for revealing well-considered self-reports*, i.e., self-reports that are honest, informed, reflected, and robust to trivial changes in viewpoint or context. Many procedures can help to create these "favourable conditions". For example, one can provide general information and/or other aids for decision-making (e.g., impartial advice from an expert); introduce truth-telling-commitment devices that help the elicitation of honest self-reports (e.g., asking people to sign a solemn truth-telling oath before giving their answers, as in Jacquemet et al. 2019, 2020); present several frames of the same issue to counter framing effects (as, e.g., in Druckman 2001, 2004; Benjamin et al. 2020); point out choices/self-reports that are inconsistent with compelling postulates of decision-making (e.g., point out non-transitive answers, as in Tversky 1969; see Benjamin et al. 2020: sec. V for a review); highlight intertemporal consequences if present (e.g., reminding people to consider implications for $T+1$ and onwards); and/or inform individuals about the timing of the policy introduction if the policy is not introduced at T (e.g., Maria, in the example above, could have asked at T if Norah would want herself to go to the theatre at $T+1$).

These favourable conditions can be interpreted as a particular ancillary condition. The previous discussion suggests a refined confirmed proxy that uses this particular ancillary condition to elicit well-considered confirmed choices:

Definition 2 (Well-considered confirmed choice). *Let $C(A, f)$ denote choosing from choice situation A with favourable conditions f. An individual is then said to considerately confirm at T her choice of $x \in C(A, d)$ made at $t < T$ if and only if at T she would want herself to select x if faced with $C(A, f)$ at T.*

In other words, a choice is said to be considerately confirmed whenever an individual would want herself to repeat that choice at the time of the welfare or policy evaluation if faced with the same menu under favourable conditions for revealing well-considered self-reports. While recording confirmed choices may be enough in a variety of contexts, recording well-considered confirmed choices seems desirable (or even necessary) in contexts that face the challenges mentioned in this discussion.

At this point, a reader may wonder whether the well-considered confirmed proxy is not similar to the reason-based proxy of welfare. To see that this is not the case, consider how data is used in each proxy. On the one hand, the reason-based proxy usually relies on indirect data about rational deliberation (e.g., eye-tracking studies that assess how attentive people are in different contexts) in order to determine which choices merit deference. In addition, this proxy usually relies on the behaviour of some individuals to determine when other individuals follow rational deliberation. On the other hand, the well-considered confirmed proxy demands self-reports to be well-considered when individuals are themselves asked to determine which of their choices merit deference. Therefore, these two proxies can lead to very different welfare inferences. According to the arguments in the previous section, the welfare inferences from the confirmed proxy are more reliable than those that arise from the reason-based proxy. The discussion in this section suggests that the well-considered confirmed proxy can correct for potential mistakes (as judged by individuals themselves) linked to "fast" thinking, that could in principle grant an advantage to the reason-based proxy in contexts where "slow" reasoning is important.

Concerning the context-independent proxy, its main advantage is that it is less demanding than the confirmed proxy in terms of data. Even though the confirmed proxy only requires gathering extra data at a single point in time, this can be a relevant advantage in some contexts. However, gathering the extra data for the confirmed proxy is important whenever one expects phenomena such as updating beliefs, changing preferences, ex-post regret, habituation, or other reasons for changes in subjective attitudes to be prevalent. It is also important when one expects conflicts between what people do and what they would like themselves to do (e.g., in cases of limited self-control). In addition, a well-considered confirmed proxy corrects for potential mistakes (as judged by individuals themselves) that can make the context-independent proxy unreliable, due, for example, to a lack of information or projection bias. The union of these phenomena seems non-empty for many (if not most) contexts, suggesting that the extra data required for the confirmed proxy is often necessary if the observer wants to have a more reliable proxy of welfare.

In sum, the arguments in Section 5 suggested that confirmed choices are a more reliable proxy of welfare than context-independent and reason-based choices. This holds especially for the period of the welfare or policy evaluation (period T). In this section, I have argued that this inference holds for other periods and many (if not most) contexts, particularly if — whenever deemed necessary — one adopts favourable conditions for revealing well-considered self-reports.

7 Nudges and Boosts

The arguments presented so far have implications for two influential behavioural policy programmes that are *non-incentivizing* (they do not provide monetary incentives) and *non-coercive*

(they do not forbid or impose options), nudges and boosts. I discuss each in turn.

Nudges generally consist of interventions that change the *choice architecture* (the background environment against which people make decisions) to alter people's behaviour in a predictable way without eliminating freedom of choice or significantly changing their economic incentives (Thaler and Sunstein 2008).[19] These interventions are usually intended to promote the welfare of targeted individuals *as judged by themselves* (Thaler and Sunstein 2008: ch. I).[20] Paradigmatic examples include displaying information about the socially acceptable behaviour of others, setting the default option to opt-out in pension schemes (as opposed to opt-in), and presenting healthier food "at eye level" in cafeterias, in order to direct individuals into more social, prudent, and healthier behaviours respectively.

Stated meta-choices address at least two challenges facing "nudgers" who are committed to improving the welfare of "nudgees" as actually judged by themselves. First, stated meta-choices provide a way for a nudger to infer whether a nudgee is making a bad choice as actually judged by the nudgee. This information is usually not known, and it is necessary (though not sufficient) for a nudge to respect the individual sovereignty of targeted individuals (see Sugden 2009: 371). Second, stated meta-choices provide a way for a nudger to distinguish between people with different goals and infer the distribution of goals in the population. This information is essential for steering individuals with different goals towards different "optimal" options and for avoiding nudges that are arbitrary or implement special interests (see Grüne-Yanoff and Hertwig 2016: 166). This means that stated meta-choices provide a criterion for choosing the "direction" to which nudges shall steer behaviour that all the while respects individual sovereignty and does not rely on nudgers' external judgement about what is best for others.

Nudges directed by stated meta-choices (hereafter *confirmed nudges*) can be targeted to a non-arbitrary sub-population. In the case of smoking, for example, a confirmed nudge that aims to steer people towards giving up smoking can target the sub-population of smokers who would want themselves to stop smoking. Importantly, this nudge would *not* interfere with smokers who have no desire to quit smoking. In practice, this could be implemented by creating a database of potential nudgees who would like to quit smoking.[21]

Confirmed nudges have at least one ethical, one behavioural, and one welfare advantage over

[19] See also Sunstein and Thaler (2003) and Thaler and Sunstein (2003). See Camerer et al. (2003) and Loewenstein and Ubel (2008) for similar policy programmes. The common approach of these programmes, often called *soft paternalism*, has been well received by many behavioural economists. See Sugden (2008, 2009), Grüne-Yanoff (2012), Qizilbash (2012), Gigerenzer (2015), and Fumagalli (2016) for critical reviews.

[20] It is worth noting that Sunstein and Thaler (2003) argue that individuals make 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" (p. 1162). This view is in line with informed preference theories and differs significantly from a view of individuals' welfare as actually judged by themselves. My arguments are addressed to people using nudges who are committed to improving targeted individuals' welfare as actually judged by themselves.

[21] The database of nudgees could be also restricted to individuals who would consent to being nudged to stop smoking. The consent would partially address the challenge to show that *nudgees really want to be nudged* (Sugden 2009). Evidence suggests that this could be done without significant influence on nudges' effectiveness (e.g. Loewenstein et al. 2015; Bruns et al. 2018).

traditional nudges that are not targeted to a sub-population. The ethical advantage lies in the fact that confirmed nudges are more likely than traditional nudges to *not* interfere with people that do not want to be nudged. It follows, I conjecture, that confirmed nudges are more likely to be accepted by people than nudges that are not targeted to a sub-population (see Hedlin and Sunstein 2016, Reisch and Sunstein 2016, and Arad and Rubinstein 2018 for surveys on the acceptability of nudges). The behavioural advantage lies in the reasonable assumption that an individual is more likely, *ceteris paribus*, to change her behaviour if she does not confirm it (e.g., a smoker is more likely to quit if she would want herself to quit than if she would not want herself to quit, all else being equal). It follows that confirmed nudges are more likely to be effective in steering the behaviour of targeted individuals than traditional nudges. Finally, confirmed nudges are more likely to have positive welfare implications than traditional nudges (both for most targeted individuals and on average). This is particularly the case when confirmed nudges are directed to a sub-population with similar goals. When this is the case, confirmed nudges exclude most people who would be made worse off from nudging and include people who, according to their stated meta-choices, could be made better off from nudging.

It is worth noting that Thaler and Sunstein (2008: 80, 116-7) sometimes appeal to a *New Year's resolution test* to support nudges. They ask, rhetorically, "how many people vow to smoke more cigarettes [...] in the morning next year?" Very few, we are prompt to agree. However, what this question omits is that many smokers do *not* vow to smoke fewer cigarettes in the morning next year either. Confirmed nudges are desirable because they have the potential to benefit targeted individuals (e.g., smokers who vow to smoke fewer cigarettes) while imposing *no* costs on individuals who confirm their behaviour (e.g., smokers who do not vow to smoke fewer cigarettes). Using stated meta-choices to target nudges, as advocated here, is therefore different from using the self-reports of a limited set of people to justify nudging other people as sometimes endorsed by Thaler and Sunstein (2008).

I now turn to the second behavioural policy programme that focuses on interventions that usually target individuals' skills, knowledge, and set of decision-making tools (their "heuristic repertoire") to help them to apply their existing or new set of competencies more effectively (see Grüne-Yanoff and Hertwig 2016; Hertwig and Grüne-Yanoff 2017; Grüne-Yanoff 2018). The goal of such interventions is to "boost" the decision makers' set of competencies such that they identify and reach their objectives (Grüne-Yanoff and Hertwig 2016). Examples include improving the representation of statistical information in health brochures to improve patients' understanding of different treatments (Gigerenzer et al. 2007), and providing physicians with "fast-and-frugal" decision trees for screening their patients in order to improve their performance in doing so (Jenny et al. 2013).

The arguments in this paper suggest that it may be relevant to boost individuals' ability and opportunities for reflecting upon their past behaviour. The underlying assumption is that by

being able to better appraise their behaviour, individuals will be better prepared to reach their objectives.

Several interventions can be justified on these grounds. A prominent example is (non-coercive) ex-post "cooling-off periods" that aim to encourage individuals to critically reconsider their own past decisions. Ex-post cooling-off periods have been used, for instance, on door-to-door sales in the U.S., by imposing that these sales need to be accompanied by a written statement informing the buyer of her right to rescind the purchase within three days of the transaction (Thaler and Sunstein 2008: 250). They allow people to change their minds after choosing, which can be relevant for people to make better decisions as judged by themselves.

It is worth noting that Thaler and Sunstein (2008) and other soft paternalists also support ex-post cooling-off periods. However, their justification lies in the ability of these interventions to countervail limited self-control. Therefore, Thaler and Sunstein (2008: 250) support ex-ante cooling-off periods as well, such as mandatory waiting periods before a couple can get divorced. Importantly, while ex-ante cooling-off periods shrink the opportunity set of an individual (Grüne-Yanoff 2012: 638-9, 644), non-coercive ex-post cooling-off periods enlarge the opportunity set with the option to cancel one's decision. The arguments in this paper provide support for ex-post cooling-off periods, and not for (mandatory) ex-ante cooling-off periods.

Another relevant example is the design of methods to elicit subjective attitudes (e.g., in the 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 encourage individuals to form considered subjective attitudes that are less liable to choice reversals (see, e.g., Slovic 1995: 369-70; Bleichrodt et al. 2001: 1499; Gilboa 2010). In line with the boost programme, one could create heuristics for boosting individual competencies that would supplement these opportunities for ex-post reflection. For example, Benjamin et al. (2020) implement a procedure for the elicitation of risk preferences that not only provides the opportunity for individuals to revise their choices, *but also* breaks down the independence axiom of expected utility theory into "baby steps" that are easy to understand.

8 Concluding Remarks

When data on choice behaviour is available and conflicting choice patterns are present, some observers are faced with the task of discriminating between choices for welfare or policy evaluation. I proposed confirmed choices as a reliable (though fallible) proxy of subjective welfare for welfare or policy evaluation at a given point in time. This proxy uses choices as the main ingredient combined with auxiliary data — stated meta-choices at the time of the welfare or policy evaluation — that is often available and sometimes easy to collect. According to the arguments presented in this paper, this proxy has decisive advantages over two influential proxies of welfare

based on context-independent and reason-based choices. Finally, I have also argued that stated meta-choices can usefully inform behavioural policy programmes that use nudges and boosts.

It is worth emphasising that the comparison of the three behavioural proxies of welfare made in this paper is sound under the proviso that it applies to the most common theories of subjective welfare in economics. Therefore, my analysis excludes subjective notions of welfare that are used less frequently in economics, such as those based on experiences of pleasure and pain or on the activation patterns of specific areas in the brain. Future comparisons of different proxies of welfare would also benefit from empirical studies (e.g. experiments) that aim to compare these proxies directly.

Finally, it is worth noting that, given the potential contingency of choices, preferences, and other subjective attitudes on changes in viewpoint or context, false or incomplete information, and adaptation, among other phenomena that may impair welfare inference, it can sometimes be difficult to identify credible rankings based on subjective information. Objective information, such as that concerning people's adaptation to their conditions, may be relevant in some contexts for welfare or policy judgements. Yet, we may not want to forget about subjective information altogether, but try instead to find richer and more reliable data sets that include information on both individual choices and self-reported attitudes. As argued above, this sometimes demands creating favourable conditions for revealing self-reports that are honest, informed, reflected, and robust to trivial changes in viewpoint or context. Hence, without the presumption of being able to recover stable and context-independent latent preferences in every situation, we may still be able to record choices and self-reported attitudes that are meaningful for normative analysis for a given context and time.

Acknowledgements

I am grateful to the editor and two anonymous referees for their valuable comments and suggestions that greatly improved the last version of this paper. I am also grateful to Jose Apesteguia, Salvador Barberà, Antoinette Baujard, Koen Decancq, Nicolas Gravel, Fabrice Le Lec, Stephane Lemaire, Olivier L'Harridon, Stephane Luchini, Marco Mariotti, Serena Olsaretti, Erik Schokkaert, Arthur Schram, Robert Sugden, Benoît Tarroux, Miriam Teschl, Alain Trannoy, Andrew Williams, and the participants at the ASSET's Annual Meeting 2017, at the Workshop in Individual Choice and Freedom, and at the Journées LAGV 2018 for valuable discussions, comments and suggestions on earlier versions of this paper. Finally, the financial support of the research project "ValFree" (The Value of Choice, grant No. ANR-16-CE41-0002-01) of the French National Agency for Research is gratefully acknowledged.

References

Adler, M. D. and Dolan, P. (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 Strack, F. (2014). From dual processes to multiple selves: Implications for economic behavior. *Journal of Economic Psychology*, 41:1–11.

Apesteguia, J. and Ballester, M. A. (2015). A measure of rationality and welfare. *Journal of Political Economy*, 123(6):1278–1310.

Arad, A. and Rubinstein, A. (2018). The people's perspective on libertarian-paternalistic policies. *The Journal of Law and Economics*, 61(2):311–33.

Ariely, D., Loewenstein, G., and Prelec, D. (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., Karlan, D., and Yin, W. (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., Niederle, M., and Sprenger, C. (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., Fontana, M. A., and Kimball, M. S. (2020). Reconsidering risk aversion. *NBER Working Paper Series*, WP 28007.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A. (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., Heffetz, O., Kimball, M. S., and Szembrot, N. (2014). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *The American Economic Review*, 104(9):2698–735.

Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association*, 7(2-3):267–319.

Bernheim, B. D. (2016). The good, the bad, and the ugly: A unified approach to behavioral welfare economics. *Journal of Benefit-Cost Analysis*, 7(1):12–68.

Bernheim, B. D. and Rangel, A. (2004). Addiction and cue-triggered decision processes. *The American Economic Review*, 94(5):1558–90.

Bernheim, B. D. and Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *The American Economic Review: Papers and Proceedings*, 97(2):464–70.

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

Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, 92:1787–94.

Bleichrodt, H., Pinto, J. L., and Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, 47(11):1498–514.

Bonein, A. and Denant-Boèmont, L. (2015). Self-control, commitment and peer pressure: A laboratory experiment. *Experimental Economics*, 18:543–68.

Botti, S. and McGill, A. L. (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.

Broome, J. (1978). Choice and value in economics. *Oxford Economic Papers*, 30(3):313–33.

Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Jonsson, M. L., and Rahali, B. (2018). Can nudges be transparent and yet effective? *Journal of Economic Psychology*, 65:41–59.

Burghart, D. R., Cameron, T. A., and Gerdes, G. R. (2007). Valuing publicly sponsored research projects: Risks, scenario adjustments, and inattention. *Journal of Risk and Uncertainty*, 35:77–105.

Camerer, C., Issacharoff, S., Loewenstein, G., O'Donaghue, T., and Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for 'asymmetric paternalism'. *University of Pennsylvania Law Review*, 151:1211–54.

Camerer, C. F., Loewenstein, G., and Prelec, D. (2004). Neuroeconomics: Why economics needs brains. *Scandinavian Journal of Economics*, 106:555–79.

Cerigioni, F. (2017). Stochastic choice and familiarity: Inertia and the mere exposure effect. Mimeo.

Cerigioni, F. (2020). Dual decision processes: Retrieving preferences when some choice are intuitive. *Journal of Political Economy (forthcoming)*.

Chambers, C. P. and Hayashi, T. (2012). Choice and individual welfare. *Journal of Economic Theory*, 147:1818–49.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *The American Economic Review*, 99(4):1145–77.

Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. (2004). For better or for worse: Default effects and 401 (k) savings behavior. In Wise, D. A., editor, *Perspectives on the Economics of Aging*, pages 81–126. University of Chicago Press, Chicago and London.

Cowen, T. (1993). The scope and limits of preference sovereignty. *Economics and Philosophy*, 9:253–69.

Dalton, P. S. and Ghosal, S. (2012). Decisions with endogenous frames. *Social Choice and Welfare*, 38:585–600.

Deaton, A. S., Kahneman, D., Krueger, A., Schkade, D., Schwarz, N., and Stone, A. (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., Fleurbaey, M., and Schokkaert, E. (2015). Happiness, equivalent incomes and respect for individual preferences. *Economica*, 82:1082–106.

Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, 22(1):91–101.

Druckman, J. N. (2004). Political preference formation: Competition, deliberation, and the (ir)relevance of framing effects. *American Political Science Review*, 98(4):671–86.

Ferreira, J. V. and Gravel, N. (2020). Chronologically-ordered rationalisable choice. Mimeo.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1):5–20.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40:351–401.

Fudenberg, D. and Levine, D. K. (2006). A dual self model of impulse control. *The American Economic Review*, 96:1449–76.

Fudenberg, D. and Levine, D. K. (2012). Timing and self-control. *Econometrica*, 80(1):1–42.

Fudenberg, D., Levine, D. K., and Maniadis, Z. (2012). On the robustness of anchoring effects in wtp and wta experiments. *American Economic Journal: Microeconomics*, 4(2):131–45.

Fumagalli, R. (2013). The futile search for true utility. *Economics and Philosophy*, 29:325–47.

Fumagalli, R. (2016). Decision sciences and the new case for paternalism: Three welfare-related justificatory challenges. *Social Choice and Welfare*, 47:459–80.

George, D. (1984). Meta-preferences: Reconsidering contemporary notions of free choice. *International Journal of Social Economics*, 11(3/4):92–107.

Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology*, 6:361–83.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., and Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2):53–96.

Gilboa, I. (2010). Questions in decision theory. *Annual Review of Economics*, 2(1):1–19.

Gilboa, I. and Schmeidler, D. (2001). *A Theory of Case-Based Decisions*. Cambridge University Press, Cambridge, UK.

Grüne-Yanoff, T. (2012). Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, 38:635–45.

Grüne-Yanoff, T. (2018). Boosts vs. nudges from a welfarist perspective. *Revue d'Économie Politique*, 128(2):209–24.

Grüne-Yanoff, T. and Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds & Machines*, 26:149–83.

Gul, F. and Pesendorfer, W. (2005). The revealed preference theory of changing tastes. *The Review of Economic Studies*, 72(2):429–48.

Gul, F. and Pesendorfer, W. (2008). The case for mindless economics. In Caplin, A. and Schotter, A., editors, *The Foundations of Positive and Normative Economics*, pages 3–39. Oxford University Press, New York.

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.

Hausman, D. M. (2016). On the econ within. *Journal of Economic Methodology*, 23(1):26–32.

Hausman, D. M. and McPherson, M. S. (2009). Preference satisfaction and welfare economics. *Economics and Philosophy*, 25:1–25.

Hedlin, S. and Sunstein, C. R. (2016). Does active choosing promote green energy use? experimental evidence. *Ecology Law Quarterly*, 43(1):107–42.

Hertwig, R. and Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6):973–86.

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 Stiglitz, J. E. (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., Lecouteux, G., and Sugden, R. (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.

Jacquemet, N., Luchini, S., Malezieux, A., and Shogren, J. F. (2020). Who'll stop lying under oath? empirical evidence from tax evasion games. *European Economic Review, 124, 103369.*, 124:1–14.

Jacquemet, N., Luchini, S., Rosaz, J., and Shogren, J. F. (2019). Truth telling under oath. *Management Science*, 65(1):426–38.

Jeffrey, R. C. (1974). Preferences among preferences. *The Journal of Philosophy*, 71(13):377–91.

Jenny, M. A., Pachur, T., Williams, S. L., Becker, E., and Margraf, J. (2013). Simple rules for detecting depression. *Journal of Applied Research in Memory and Cognition*, 2:149–57.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus & Giroux, New York, NY.

Kahneman, D. and Deaton, A. S. (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 Riis, J. (2005). Living, and thinking about it: Two perspectives on life. In Huppert, F. A., Kaverne, B., and Baylis, N., editors, *The Science of Well-being*, pages 285–304. Oxford University Press.

Kahneman, D., Wakker, P., and Sarin, R. (1997). Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics*, 112:375–406.

Kirman, A. and Teschl, M. (2006). Searching for identity in the capability space. *Journal of Economic Methodology*, 13(3):299–325.

Koszegi, B. and Rabin, M. (2007). Mistakes in choice-based welfare analysis. *American Economic Review Papers and Proceedings*, 97(2):477–81.

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.

Little, I. M. D. (1949). A reformulation of the theory of consumer's behaviour. *Oxford Economic Papers*, 1(1):90–9.

Loewenstein, G., Bryce, C., Hagmann, D., and Rajpal, S. (2015). Warning: You are about to be nudged. *Behavioral Science and Policy*, 1(1):35–42.

Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics*, 118(4):1209–48.

Loewenstein, G. and Thaler, R. H. (1989). Anomalies: Intertemporal choice. *Journal of Economic perspectives*, 3(4):181–93.

Loewenstein, G. and Ubel, P. A. (2008). Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics, 92(8-9), 1795-1810.*, 92:1795–810.

Maniadis, Z., Tufano, F., and List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *The American Economic Review*, 104(2):277–90.

Manzini, P. and Mariotti, M. (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.

Nussbaum, M. (2006). *Frontiers of Justice: Disability, Nationality, Species Membership*. Harvard University Press, Cambridge, MA.

OECD (2017). Behavioural insights and public policy: Lessons from around the world. http://dx.doi.org/10.1787/9789264270480-en, OECD Publishing, Paris.

Office for National Statistics (2019). Adult smoking habits in the uk 2019. `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2019`.

Olsaretti, S. (2006). Introduction. In Olsaretti, S., editor, *Preferences and Well-being*, pages 1–7. Cambridge University Press, Cambridge, UK.

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

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. and van Leeuwen, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational Behavior and Human Decision Processes*, 76(2):189–205.

Reisch, L. A. and Sunstein, C. R. (2016). Do europeans like nudges? *Judgment and Decision Making*, 11(4):310–25.

Rosati, C. (1995). Persons, perspectives, and full information accounts of the good. *Ethics*, 105:296–325.

Rubinstein, A. and Salant, Y. (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 Rubinstein, A. (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.

Sen, A. K. (1973). Behaviour and the concept of preference. *Economica*, 40(159):241–59.

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.

Sobel, D. (2009). Subjectivism and idealization. *Ethics*, 119(2):336–52.

Sugden, R. (2004). The opportunity criterion: Consumer sovereignty without the assumption of coherent preferences. *The American Economic Review*, 94(4):1014–33.

Sugden, R. (2008). Why incoherent preferences do not justify paternalism. *Constitutional Political Economy*, 19(3):226–48.

Sugden, R. (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):365–73.

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

Sunstein, C. R. and Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4):1159–202.

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

Thaler, R. H. and Sunstein, C. R. (2003). Libertarian paternalism. *The American Economic Review*, 93:175–9.

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

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31–48.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–31.

Velleman, J. D. (1999). A right to self-termination? *Ethics*, 109:606–28.

Waldfogel, J. (2005). Does consumer irrationality trump consumer sovereignty? *Review of Economics and Statistics*, 87(4):691–96.