ANOMALIES OR EXPECTED BEHAVIORS? UNDERSTANDING STATED PREFERENCES AND WELFARE IMPLICATIONS IN LIGHT OF CONTEMPORARY BEHAVIORAL THEORY

LEONHARD K. LADES
EWA ZAWOJSKA
ROBERT J. JOHNSTON
NICK HANLEY
LIAM DELANEY
MIKOŁAJ CZAJKOWSKI

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS

Anomalies or Expected Behaviors? Understanding Stated Preferences and Welfare Implications in Light of Contemporary Behavioral Theory

Leonhard K. Lades[a], Ewa Zawojska[b], Robert J. Johnston[c], Nick Hanley[d], Liam Delaney[e], Mikołaj Czajkowski[b]

[a] *Environmental Policy & Geary Institute, University College Dublin*
[b] *University of Warsaw, Faculty of Economic Sciences*
[c] *George Perkins Marsh Institute and Department of Economics, Clark University*
[d] *Institute of Biodiversity Animal Health & Comparative Medicine, University of Glasgow*
[e] *Department of Psychological and Behavioural Science, The London School of Economics and Political Science*

Corresponding authors: leonhard.lades@ucd.ie, ewa.zawojska@uw.edu.pl, rjohnston@clarku.edu, Nicholas.Hanley@glasgow.ac.uk, L.D.Delaney@lse.ac.uk, mc@uw.edu.pl

Abstract: The stated preference literature contains an expansive body of research on behavioral anomalies, typically understood as response patterns that are inconsistent with standard neoclassical choice theory. While the literature often implies that anomalous behaviors are distinct to stated preferences, widespread evidence of similar patterns across real-world settings raises the potential for an alternative interpretation. We argue that these anomalies might actually reflect behaviors that are to be expected once deviations from the standard economic model and behavioral reactions to the choice architecture in stated preference surveys are considered. The article reviews and organizes the evidence of so-called "anomalous" stated preference behaviors within the context of behavioral science to provide guidance for applied welfare economics. We coordinate evidence on these anomalies using a typology grounded in behavioral science, which groups non-standard behaviors into: non-standard preferences, non-standard beliefs, and non-standard decision-making. We apply this typology to organize the evidence, clarify nomenclature, and understand the implications of non-standard behaviors in stated preference studies for applied welfare analysis. Observing the systematic and common nature of these behaviors in actual and hypothetical settings, we outline possibilities to overcome associated challenges for applied welfare analysis, by adapting new frameworks for welfare analysis proposed within behavioral science.

1. Introduction

The stated preference literature contains an expansive body of research on behavioral anomalies (Johnston, Boyle, et al., 2017; Kling, Phaneuf, & Zhao, 2012; Poe, 2016). These anomalies are typically characterized as systematic response patterns that are inconsistent with *a priori*, theoretical predictions drawn from traditional neoclassical theory of choice. Although this literature often implies that such anomalous behaviors are distinct to stated preferences, there is widespread and increasing evidence from behavioral science that similar deviations from traditional neoclassical theory appear across a wide range of decision contexts, from markets to lab experiments to field survey studies (DellaVigna, 2009; Thaler, 2015).

This article reviews and organizes the evidence of so-called "anomalous" stated preference behaviors within the context of contemporary behavioral science research that seeks to improve the realism of the traditional economic model and provide guidance for applied welfare economics. In contrast to the common interpretation of anomalous behaviors as evidence of invalidity in stated preference studies, behavioral science provides an alternative interpretation of the same behaviors as predictable expressions of non-standard preferences, non-standard beliefs, and/or non-standard decision-making that emerge due to consequences of the situations—also called choice architecture—within which people make decisions. From this perspective, stated preference anomalies might actually reflect behaviors that are to be expected once deviations from the standard economic model and behavioral reactions to the choice architecture are considered.[1] This conceptualization of anomalous behavior may enable new approaches to welfare analysis that accommodate these behaviors in a systematic manner. Interestingly and perhaps promisingly, these approaches share many common themes with current best practices in stated preference welfare analysis.

The first goal of the paper is a systematic understanding of anomalies observed in stated preference studies that is consistent with emerging theory and findings from behavioral science. Within the behavioral economics literature, the predictability of non-standard, systematic response patterns has been enhanced by the organization of these behaviors around three primary deviations from the traditional economic model: non-standard preferences, non-standard beliefs, and non-standard decision-making (DellaVigna, 2009). Based on this

---

[1] Similar observations have been made before (e.g., Carson, 2012; Haab et al., 2013; Kahneman et al., 1999; Kling et al., 2012). For example, Poe (2016) presents seven stated preference anomalies that have also been observed in actual choice settings. However, we extend these prior contributions by organizing the anomalies according to current behavioral science typology and identifying implications for welfare analysis based on this perspective.

structure, we coordinate evidence from the literature to propose a constructive perspective on how and why non-standard behaviors manifest in stated preference studies, and how these behaviors parallel those observed in non-hypothetical choice contexts. This typology provides a structure through which to clarify nomenclature and organize evidence of non-standard behaviors observed within the stated preference literature. In doing so, we distinguish between stated preference anomalies and their behavioral foundations, and apply the latter to propose a framework to organize prior theoretical and empirical findings on these anomalies.

The second goal is to derive implications for applied welfare analysis using stated preference data. The systematic nature of the non-standard behaviors in both stated and revealed preferences implies possibilities to overcome challenges for applied welfare analysis caused by anomalous behaviors. Once one acknowledges that many behaviors, including those in stated preference settings, are influenced by similar behavioral foundations and effects of the choice architecture, it becomes possible to consider more holistic approaches to non-standard behaviors within all types of welfare economics, rather than imposing arguably arbitrary distinctions between stated and revealed preferences. We therefore consider how innovative frameworks for welfare analysis (e.g., benefit-cost analysis) proposed within behavioral science might be applicable to environmental economics and the future research that would be required. In concept, insights from a new "behavioral welfare economics" could be used to develop practical guidance for what non-standard behaviors imply for welfare analysis based on stated preference data.

2. Organizing stated preference "anomalies"

Research in behavioral economics aims to improve the psychological realism of the traditional economic model. To do so, findings are frequently organized according to three systematic deviations from the standard economic model: non-standard preferences, non-standard beliefs, and non-standard decision-making (DellaVigna, 2009; Rabin, 1998). Primary *non-standard preferences* discussed in behavioral economics include: social preferences, risk preferences, and time preferences. *Non-standard beliefs* are captured in the literature on three broad domains: incorrect predictions about oneself in the future, the representativeness heuristic (i.e., the likelihood of outcomes being perceived based on how representative those outcomes are), and the tendency to obtain utility directly from beliefs. Finally, one can differentiate between four main types of *non-standard decision-making*: limited attention, limited cognitive effort, social influences, and effects of emotions.

In the last decade, interactions between scholars at the interface of economics, psychology, and cognate disciplines have broadened the analysis and started to investigate how these three deviations interact with characteristics of the choice environment (or architecture) to influence behavior and welfare. Sometimes called "supposedly irrelevant factors", these characteristics (e.g., defaults, reminders, and labels) have been shown to have strong and systematic effects on behavior (Thaler, 2015). At least four groups of context factors have been identified: defaults, simplification and salience, psychologically informed messaging, and sludge.

The Supplementary Material to this article presents a more detailed explanation of the key insights from the behavioral economics literature on the three types of deviations from the standard economic model and the influence of the choice architecture on these deviations. We focus therein on insights that are particularly salient for our organization of stated preference anomalies. More detailed summaries of the underlying behavioral science and applications to other areas of research are presented elsewhere (Benjamin, 2019; Carlsson, Gravert, Johansson-Stenman, & Kurz, 2021; DellaVigna, 2009; Dhami, 2016; Kremer, Rao, & Schilbach, 2019).

This section uses the structure suggested by behavioral economics to organize the literature on stated preference anomalies. We make a distinction between stated preference anomalies and the behavioral explanations for these anomalies, and use the latter to organize the literature. To make our point as clearly as possible—that anomalies are not really anomalies but expected effects—we assume here that the stated preference literature has identified "effects," but has not explained these effects systematically within a framework that is consistent with behavioral science. While this is a bit of a strawman argument, it is helpful to distinguish "effects" and "behavioral explanations" for these effects, particularly as one moves forward to consider implications for welfare analysis. The section further builds on recent work showing that many stated preference anomalies can be mapped to similar behaviors in other contexts such as incentivized experiments and field studies (Carson, 2012; Haab, Interis, Petrolia, & Whitehead, 2013; Kahneman, Ritov, & Schkade, 1999; Kling et al., 2012; Poe, 2016), but extends this work by organizing the literature according to the three categories of deviations noted above. For each category, some of the most common stated preference anomalies are discussed.[2]

---

[2] In the interest of conciseness, we do not provide an encyclopedic list or review of all possible anomalies discussed in the stated preference literature—although most if not all of these may be considered using the framework developed here.

2.1. Non-standard preferences

2.1.1. Excess disparity between willingness to pay (WTP) and willingness to accept (WTA)

The excess disparity between WTP and WTA measures is a commonly discussed anomaly in stated preference literature. Survey participants often demand (much) higher compensations for a loss of a good (WTA) compared to what they would pay to for the same good (WTP) (Brown & Gregory, 1999; Horowitz & McConnell, 2002; Tunçel & Hammitt, 2014). Non-standard risk preferences, and more specifically prospect theory, suggests one explanation for this disparity (Kahneman & Tversky, 1979; Thaler, 1980).[3] Participants may have reference-dependent preferences and evaluate *any* prospect in comparison to a reference point. In WTP elicitation, this reference point is *not* having the good and in WTA elicitation, it is having the good. Prospect theory suggests that people are more averse to losses than they are keen to acquire gains, and this loss aversion makes WTA values larger than WTP measures. These effects are not unique to stated preferences. Examples of prospect theory in the field are available from many contexts such as consumption (people are more sensitive to price increases than to cuts), macroeconomics (consumers do not reduce consumption after bad income news), the stock markets (stock returns are too high relative to bond returns), and the labor market (cab drivers in New York work until they reached their individually-set daily target) as summarized in recent work (Barberis, 2013; Camerer, 2004). Further description of the WTP/WTA disparity, including additional explanations is provided by Poe (2016).

2.1.2. Unselfish preferences

Some stated preference studies find that participants consider the welfare of others as well as their own when indicating their valuations of goods.[4] People's value orientations (e.g., altruism; McConnell 1997) are associated with variations in WTP (Ojea & Loureiro, 2007; Poe, 2016). Behavioral theories which incorporate fairness, altruism and other social preferences into the utility function can explain these effects (e.g., Andreoni 1990; Fehr and Schmidt 1999). Behavior motivated by unselfish preferences is widespread in domains such as charitable giving (DellaVigna, List, & Malmendier, 2012) and the motivation to work (Gneezy & List, 2006).

---

[3] Other explanations include income effects and substitutes, transaction costs, responsibility, and others (Brown & Gregory, 1999).

[4] Unselfish preferences are not always described as an anomaly, but they do imply a challenge for benefit-cost analysis (Johansson, 1992).

2.1.3. Insensitivity to scope

Answers in stated preference surveys do not always differ as expected when the scope of the good (its quantity or quality) is changed (Carson, 1997). For example, an improvement on 1,000 acres of a forest is not always valued significantly higher than a similar improvement on 500 acres of the forest. A related concern is welfare estimates that do not meet the "adding-up" property, which implies that the value of a bundle of outcomes should be equal to the sum of the values of the constituent parts (Bishop & Boyle, 2021). There are multiple explanations for a lack of scope sensitivity and adding-up that are consistent with the standard economic model (Whitehead, 2016), and behaviors of this type are not unique to stated preferences (Bishop & Boyle, 2021; Poe, 2016). One possible behavioral economic explanation related to social preferences is the desire to obtain moral satisfaction or warm glow from one's answer in the stated preference study (Kahneman & Knetsch, 1992). Other explanations include mental accounting and reciprocity—the latter being a form of social preference discussed above (Bishop & Boyle, 2021; Thaler, 1999).

2.2. Non-standard beliefs

2.2.1. Social-image and self-image considerations

Answers to stated preference surveys can be influenced by participants' preferences to appear in a positive light in front of others and in front of themselves (Börger, 2013; Lusk & Norwood, 2009). Behavior of this type has long been recognized in settings such as experiments and surveys, and is often called social desirability bias (Nederhof, 1985). Models of belief-based utility can explain that these beliefs about one's image are part of the utility function and thus influence stated preferences. Image concerns have been shown to explain field behaviors in contexts such as take-up of social transfers (Friedrichsen, König, & Schmacker, 2018) and blood donation (Lacetera & Macis, 2010).

2.2.2. Temporal reliability of stated preferences

Temporal reliability of stated preferences describes the stability of measures over time and context (Mitchell and Carson, 1989; Carson et al, 1997). Of particular concern is instability that manifests over short time periods, during which (absent an external shock of some type) economists generally assume that preferences should be stable. Temporal reliability can be shown by comparing answers given by the same individuals at different points in time to the

same survey or by comparing the outcomes of similar studies across different samples conducted at different times. For example, He & Zhang (2021) show that the WTP for air quality improvements depends on the air quality on the day on which their contingent valuation surveys are administered. One explanation for this effect is projection bias, which is a non-standard belief that has been documented in the field in contexts such as the effect of weather on people's consumption choices (Busse, Pope, Pope, & Silva-Risso, 2015; Conlin, O'Donoghue, & Vogelsang, 2007) and on voting decisions (Meier, Schmid, & Stutzer, 2019).

2.3. Non-standard decision-making

2.3.1. Starting point bias

Starting point bias describes the influence that the presentation of an initial value or situation has on subsequent decisions (Boyle, Bishop, & Welsh, 1985; Herriges & Shogren, 1996). For example, the bid value or environmental improvements offered in initial choice questions may systematically influence responses to subsequent valuation questions. One explanation of starting point bias that is discussed as non-standard decision-making in behavioral economics is anchoring (Herriges & Shogren, 1996; McFadden, 1994). Anchoring describes the effect that the initial exposure to a number can have on subsequent judgments (Tversky & Kahneman, 1974). Initially observing a certain number (e.g., a household cost or scope of an environmental change) can create a mental reference point. Examples of anchoring can be found in many real-world situations. For example, Sakaguchi et al. (2022) suggest that presenting credit card debtors with the minimum payments to pay down their debt as a default anchors the manual repayment decision on the minimum payment to the minimum. As a result, they suggest that approximately eight percent of all of the interest ever paid in the UK is due to this effect.

2.3.2. Sequencing effects

Sequencing effects are defined as situations in which stated preference responses are sensitive to the order in which questions are presented (Day et al., 2012). Starting point bias is one particular type of sequencing effect and thus anchoring is one of the behavioral explanations. More general sequencing effects can be explained by behavioral research on query theory, which describes one form of non-standard decision-making. Query theory proposes that people process queries one after the other rather than simultaneously, so that earlier queries receive more thought than later queries and are thus more influential (Johnson, Häubl, & Keinan, 2007). Examples of sequencing effects in real-world settings include food choice (Gravert &

Kurz, 2021) and the link between ballot order and election outcomes (Ho & Imai, 2008; Miller & Krosnick, 1998).

2.3.3. Attribute non-attendance

Attribute non-attendance describes selective (in)attention wherein a respondent ignores one or more aspects of the scenario description in a stated preference study (Alemu, Mørkbak, Olsen, & Jensen, 2013; Campbell, Hensher, & Scarpa, 2011; Scarpa, Gilbride, Campbell, & Hensher, 2009). For example, in stated preference discrete choice experiments, respondents may ignore some attributes. Behavioral economic theories on limited attention formalize situations in which decision makers do not consider all relevant attributes (Allcott & Taubinsky, 2015; Gabaix, 2019). Attribute non-attendance has been identified in the real world as well, for example, when consumers pay only partial attention to taxes that are added later to the price of a good (Chetty, Looney, & Kroft, 2009), or ignore "shrouded attributes" such as "surprise" bank fees or add-ons when booking flights (Ellison & Ellison, 2009; Gabaix & Laibson, 2006).

2.3.4. Yea and nay saying

Yea and nay saying is present when respondents agree or disagree to proposed policy scenarios without consideration of their characteristics, including the price charged. One reason for this behavior is that respondents believe that the policy has been designed by someone for a good reason; this has been discussed as one reason for the stickiness of defaults (Carlsson, 2010). Default effects are arguably the finding with the strongest effect size within behavioral economics in experiments and the field (Jachimowicz, Duncan, Weber, & Johnson, 2019). These behaviors can also be framed as extreme examples of attribute non-attendance (where most or all attributes are ignored) and in the case of nay saying can reflect status quo bias—hence the behavioral explanations for these phenomena apply here as well.

3. Implications of behavioral welfare economics for stated preference welfare analysis

Stated preference methods are often used to support applied welfare analysis (Johnston et al. 2017). This analysis relies on the assumption that people make rational, compensatory decisions that maximize their welfare.[5] If people make rational decisions within an incentive-

---

[5] The standard economic welfare calculus relies on three premises: that each of us is the best judge of our own well-being; that our judgments are governed by coherent, stable preferences; and that our preferences guide our choices—when we choose, we seek to benefit ourselves (Bernheim, 2016; Hausman, 2012).

compatible setting, their choices are assumed to reveal their true preferences and money-metric values (often measured in terms of WTP).[6] Welfare measures of this type are suitable to guide welfare calculations in normative policy analysis, such as benefit-cost analysis (e.g., Hausman, 2012). Both the stated preference and behavioral economics literatures acknowledge that systematic deviations from rational decision making (DellaVigna, 2009) can pose challenges for all economic welfare analysis (e.g., Bernheim, 2016; Sunstein, 2020a). However, unlike the emphasis of much of the stated preference literature, which tends to frame these concerns as biases and anomalies that may jeopardize validity, behavioral welfare economics has taken a more constructive approach, considering how welfare analysis can proceed given that systematic deviations from rational decision making are common across many decision settings.

Interpreting stated preference anomalies from the perspective of behavioral economics provides the opportunity to consider whether and how insights from behavioral welfare economics might apply to stated preferences. In some instances, these approaches imply fundamental differences in how stated preference methods are approached or interpreted, while in others they imply minor adjustments to longstanding techniques used for stated preference design and analysis. The next three subsections summarize the main behavioral critiques of standard welfare economics (subsection 3.1), recap the main strategies for redesigning welfare economics as suggested by behavioral welfare economics (subsection 3.2), and present recommendations for stated preference studies based on these strategies (subsection 3.3).

3.1. Behavioral economic critiques of standard welfare economics

There are two key behavioral economic critiques of standard welfare economics: the "implementation critique" and the "coherence critique" (Bernheim, 2016, 2021). The implementation critique argues that people may have difficulties making choices that advance their own objectives so that the decision utility which informs choice is not the same as the experienced utility that is welfare relevant (Beshears, Choi, Laibson, & Madrian, 2008; Kahneman, Wakker, & Sarin, 1997). The critique suggests that people do have true preferences, but that their decisions do not always correctly "implement" these preferences. Implementation failures can occur when people's decisions rely on incorrect assumptions about

---

[6] Although WTA is a more conceptually appropriate welfare measure in some contexts, WTP is often selected due to practical challenges associated with WTA estimation (e.g., difficulty framing incentive-compatible questions and assuring perceived credibility of monetary compensation; Johnston et al. 2017; Lloyd-Smith & Adamowicz 2018).

which options are available, when people have mistaken beliefs about the consequences of their actions, or when people's decisions are systematically biased relative to their true underlying (experienced) utility. We have mentioned many examples for such implementation failures in section 2. For example, behaviors such as attribute non-attendance, yea or nay saying, and scope insensitivity can be interpreted as examples of implementation failures. The failure of observed choices to reveal underlying experienced utility within non-incentive-compatible choice settings can also be interpreted in this manner (although see Börger et al. 2022).

The coherence critique to standard welfare economics is even more problematic for welfare analysis. It suggests that true preferences that are coherent and stable do not exist in the first place because preferences are constructed dependent on the context in which they are expressed (Bernheim, 2016). In other words, the critique suggests that there is no "inner rational agent" with neoclassical preferences (Infante, Lecouteux, & Sugden, 2016). If coherent preferences do not exist, they cannot be revealed by choice data or in stated preference surveys (nor can they be revealed by any type of behavior). The critique relies on psychological findings suggesting that preferences are dependent on the frames in which the decisions are presented (Chater, 2018; Dold, 2018; Johnson et al., 2007; Kahneman & Tversky, 1984; Lichtenstein & Slovic, 2006; Sunstein, 2020a). Different frames lead to different forms of preferences and preferences without frames do not exist. The frames, however, are normatively irrelevant. Behaviors such as anchoring and coherent arbitrariness (Ariely, Loewenstein, & Prelec, 2003; Hanley, Kriström, & Shogren, 2009) and the influence of normatively irrelevant factors, such as the weather on outcomes such as car purchases (Busse et al. 2015) or voting (Meier et al. 2016), can be used as evidence that coherent preferences do not exist.[7]

3.2. Behavioral economic welfare analysis

Whilst many approaches have been suggested to deal with the problems posed by behavioral insights for economic welfare analysis, here we emphasize choice-based welfare approaches that start from the assumption that welfare is closely related to the satisfaction of preferences, which can be revealed in choice data. This is the closest and hence most applicable line of approaches to what is typically done in stated and revealed preference valuation studies. Below we summarize the most common choice-based approaches to derive valid normative welfare

---

[7] In some cases, implementation failures and coherence failures may be observationally equivalent. For example, anchoring may be due to an underlying lack of coherent preferences or may be due to mistaken beliefs generated during earlier choices that contribute to later implementation failures. Hence, it may not always be straightforward to identify and disentangle the underlying causes of apparent anomalous behavior, either in stated or revealed choices.

estimates, as characterized in the behavioral economics literature, and discuss how these are (or are not) parallel to current best practices for stated preference studies.

3.2.1. *Purifying preferences*

One influential approach suggests that it is possible to identify preferences that can be interpreted as true preferences even when decisions are influenced by anomalies or biases (Beshears et al., 2008; Sunstein, 2020a). This approach aims to identify the decisions that individuals *would have made* in the absence of psychological or informational distortions (Chetty et al., 2009; Thaler & Sunstein, 2008). To do so, the welfare analyst's task is to identify the influence of assumed or confirmed biases on decision-making and then analytically "reconstruct" the normative preferences, so removing the influence of these biases. If enough is known about the structural influence of biases on decisions (or conditional upon assumptions about these influences), both preference parameters and bias parameters can be recovered simultaneously from choice data. This process is sometimes called "preference purification" (Hausman, 2012). While it acknowledges the implementation critique (i.e., that choices do not always implement true preferences), it does not deal with the coherence critique and assumes that coherent and stable preferences do exist (Infante et al., 2016). Moreover, the approach is difficult (and sometimes impossible) to apply in practice because the analysis requires one to identify which aspects of a choice reflects a bias versus a true underlying preference (Bernheim, 2021).

3.2.2. *Referring to choices made under epistemically favorable conditions*

A second approach is to restrict the choices that can be interpreted as revealing true preferences to those that are "adequately informed and sufficiently free from behavioral biases" (Beshears et al., 2008; Sunstein, 2020a). This approach does not require the identification of specific bias parameters to purify preferences. Instead, it is sufficient to identify choices that can be assumed to be free of mistakes and then prioritize these choices in welfare analyses (Sunstein, 2020a, 2020b).

Welfare-relevant choices can be identified based on several characteristics: (1) choices unaffected by clearly irrelevant frames, (2) choices made with complete information, (3) choices made actively, (4) choices not influenced by decision making biases, and (5) choices that are made when people consider the big picture without suffering from limited attention (Sunstein, 2020b). Compared to preference purification, this approach is easier to implement

in practice. However, it is still subject to the coherence critique, and inconsistent choices are ignored in the welfare calculation although they might be normatively meaningful.

3.2.3. *Bernheim-Rangel framework*

The Bernheim-Rangel approach was developed partly to accommodate both the implementation critique and the coherence critique (Bernheim, 2016, 2021; Bernheim & Rangel, 2009; Bernheim & Taubinsky, 2018). The framework does not assume that people's decisions always implement their true preferences (because people make mistakes), nor does it assume that an inner rational agent with normative preferences exists (because preferences can be contextually constructed). It does, however, suggest that there is normative information even in inconsistent choices if these are not made mistakenly. The approach allows for some ambiguity to accommodate normatively relevant, contextually constructed preferences.[8]

In practice, the framework includes two key tasks. The first task is to identify the subset of choices that merits deference for welfare analysis. These choices are part of the "welfare relevant domain" and include all choices that cannot be classified as mistakes. Mistaken choices are not to be used for welfare analysis. To identify mistakes in practice, the analysist first needs to demonstrate that a behavior is context-dependent as there is no room for the welfare analyst to claim that a mistake exists without context dependence (even if the analysis thinks that the choice is not welfare improving).[9] Once context dependence has been established, the analysist tests whether this context dependence is the result of an incorrect characterization of the available opportunities, i.e., whether it is a "characterization failure" and not just the result of contextually constructed preferences. To identify characterization failures, the analysis evaluates whether the choice is the result of a misunderstanding of the available options and consequences, conditional on the decision maker's available information. The welfare relevant domain excludes all choices that are thus identified as mistaken, but it can include choices that are based on constructed preferences, so that inconsistencies across choices within the welfare relevant domain are possible.

---

[8] In a clarifying illustration, Bernheim (2021) states that "we do not, for example, dismiss the opinions of a panel of experts simply because they do not see eye to eye on every point." Similarly, he argues, one should not dismiss choices if they are not totally coherent with other choices.

[9] This claim has been criticized, for example, by Sunstein (2020a) who suggests that even context-independent choices can be mistaken because people sometimes make consistent mistakes about what choices make their lives go well, particularly when they must predict future welfare effects.

Once the welfare relevant choices have been identified, the second task is to use the choices in this domain to infer whether one option provides more welfare than another. The framework suggests that one good *x* is preferred to another good *y* if the welfare relevant domain contains no decision problem in which *x* is available where the consumer is willing to choose *y*. It also suggests that when there are choice inconsistencies, there is ambiguity in the preferences. In practice, this means that WTP (or WTA) measures which differ across decision frames within the welfare relevant domain provide lower and upper bounds of people's true WTP (or WTA). This acknowledged ambiguity in the elicited value acknowledges that preferences can be contextually constructed and still be normatively valid (as long as the choices are from the normatively relevant domain and not are not made mistakenly). This approach then allows for applying concepts typically used in economic welfare economics, such as compensating and equivalent surplus.

3.3. *What does this imply for stated preference studies?*

Given that non-standard preferences, non-standard beliefs, and non-standard decision-making can be identified in both stated and revealed preferences, the implementation critique and the coherence critique apply to stated preference studies as well. As such, at least in concept, welfare analysis using stated preference studies can incorporate any of the three approaches developed by behavioral scientists in the context of revealed preferences to address these two critiques. The following provides a discussion on how the three approaches of behavioral welfare analysis might be applied to stated preference studies.

3.3.1. *Purifying stated preferences*

Stated preference studies were historically among the earliest studies to identify presumed biases in decision-making and welfare estimates (e.g., Davis 1963; Bishop and Heberlein 1979; Carlsson, 2010). The literature includes a long tradition of studies that have sought to identify the extent of bias (or lack thereof) in stated preference welfare estimates and to calibrate these estimates to provide more valid estimates of "true" welfare (e.g., Johnston 2006; Johnston et al. 2017; List and Gallet 2001; Murphy et al. 2005). For example, some analysts have proposed certainty adjustments (Champ, Moore, & Bishop, 2009), assessed the magnitude of bias in hypothetical value estimates (e.g., List and Gallet 2001; Little and Berrens 2004; Penn and Hu 2018), or suggested the use of survey components such as cheap talk (Cummings & Taylor, 1999) or oath scripts (Jacquemet, Joule, Luchini, & Shogren, 2013) in an attempt to purge

welfare estimates of presumed hypothetical biases. Approaches of this type often require strong assumptions regarding the structure of preferences and/or the biases that occur, and are not generally recommended as best practices (Johnston, Boyle, et al., 2017).

In other cases, structural assumptions have been made in an attempt to reconstruct what estimated preferences *might have been*, were respondents to have made choices under settings that did not induce suspected biases—for example, if survey respondents were provided with information that was more consistent with their underlying preferences (Johnston, Schultz, Segerson, Besedin, & Ramachandran, 2013, 2017).[10] As a final example, the attribute non-attendance literature has proposed a variety of empirical methods to correct for and offset attendant effects on welfare estimates (Campbell et al., 2011; Lew & Whitehead, 2020; Scarpa et al., 2009), and thereby (at least in a sense) "purify" the resulting preference information. Note that all approaches of this type require at least some maintained (if often implicit) assumptions to distinguish biases or anomalies from true preferences.

Grounded in this tradition, it might be possible to design stated preference studies in a way that allows one to structurally estimate bias parameters as well as preference parameters from stated choices. As noted above, multiple areas of the stated preference literature are grounded in the concept that true underlying preferences can in principle be recovered by identifying and purging undesirable biases or anomalies (Alemu et al., 2013; Campbell et al., 2011; Lew & Whitehead, 2020; Lloyd-Smith & Adamowicz, 2018; Scarpa et al., 2009; Vossler, Doyon, & Rondeau, 2012; Vossler & Zawojska, 2020). As such, preference purification shares many similarities with (sometimes but not always ad hoc) approaches to adjust or calibrate stated preference welfare measures to better reflect "true" underlying preferences.

This work notwithstanding, the stated preference literature has not yet developed a cohesive theoretical framework to inform preference "purification" of this type from a holistic standpoint. Existing examples target individual anomalies and biases. Moreover, these typically assume that revealed preferences or behavior in economic lab experiments provide true measures of underlying preferences (supporting *criterion* validity tests), rather than treating them more appropriately as alternative measures of the same underlying construct (thereby supporting *convergent* validity tests) (Bishop & Boyle, 2019; Johnston, Boyle, et al.,

---

[10] Johnston et al. (2013; 2017) consider cases where evaluated policy scenarios omit information on an ecological attribute (or outcome) that is relevant to respondents but include other, biophysically related information. They assume simple structural models in an attempt to infer whether and how respondents use the *included* information to draw (perhaps erroneous) conclusions about the *omitted* outcome, and what welfare estimates would have been if respondents had received information on the omitted (but nonetheless valued) outcome.

2017). Other approaches make theoretically questionable assumptions, for example that reductions in estimated WTP are sufficient evidence that results are more valid (and hence that the approach is "successful"). Hence, significant progress will be required to move from the current state of the literature to a holistic framework for preference purification within stated preferences.

Similarly, behavioral welfare economists have not yet made much progress in designing efficient approaches to cleanly distinguish between preferences and biases in choices (Bernheim, 2021). As such, neither the behavioral economics nor stated preference literatures provide sufficient foundation today for the estimation of purified preferences in stated preference studies that offset all possible deviations from standard neoclassical assumptions. However, stated preference research can make progress in this regard. An advantage of stated preference studies is that the study materials are designed by the analyst, and thus it should be easier to design studies that are able to differentiate between bias and preferences in the decisions compared to field studies. To this end, there are advantages of integrating the insights and models from both strands of literature (behavioral economics and stated preferences) to address whether and how more universal preference purification might be possible.

3.3.2. *Referring to stated choices made under epistemically favorable conditions*

Applying this second behavioral welfare economics approach to stated preference research suggests that researchers should be more confident in the validity of their data when respondents in the surveys, for example, make consistent choices, are well-informed, make active choices rather than sticking to or always rejecting the status quo option, are not influenced by known decision making biases, and make their choices with a broad perspective. As a simple example, if one is concerned with anchoring effects on subsequent valuation questions in a sequence (assuming these are due to implementation rather than coherence failures), one might use responses only from one-shot (single question) stated preference value elicitations for welfare analysis (or perhaps from only the first question in a sequence), under the assumption that the resulting estimates will be free of these unintended effects. Similarly, concerns with attribute non-attendance or lexicographic preferences might be addressed by providing respondents with simpler choice tasks.

In many ways, existing guidelines for stated preference research reflect a similar goal: to develop survey questions and scenarios that provide epistemically favorable conditions for welfare elicitation (Johnston, Boyle, et al., 2017). For example, the heavy emphasis on good

survey design and pretesting (including focus groups, cognitive interviews, and supporting techniques such as verbal protocols) encourages the development of survey instruments in which responses that support valid and reliable welfare estimation can be elicited (Bishop & Boyle, 2019; Carson, 2012; Johnston, Boyle, et al., 2017; Johnston, Weaver, Smith, & Swallow, 1995; Kaplowitz, Hadlock, & Levine, 2004; Schkade & Payne, 1994). Similar motivations apply to the use of incentive-compatible questions within welfare elicitation (Carson & Groves, 2007; Vossler et al., 2012; Vossler & Zawojska, 2020) and guidance to limit the complexity of evaluated policy scenarios (DeShazo & Fermo, 2002; Swait & Adamowicz, 2001), among other recommended practices in stated preference survey design (Johnston et al. 2017). As another illustrative example, past work has sought to identify effects of respondent experience on response behavior, grounded in the assumption that more experienced and knowledgeable respondents (or those whose lack of experience is offset by survey design approaches such as virtual reality) are more likely to provide welfare-revealing survey responses (Cameron & Englin, 1997; Carlsson, 2010; LaRiviere et al., 2014).

Indeed, one might argue that much of the contemporary stated preference literature is predicated on the viability of designing epistemically favorable conditions for welfare elicitation. However, thus far approaches such as these have not been developed within (or informed by) the formal frameworks proposed by behavioral welfare economists for recovery of true preferences. Instead, they have developed "organically" in the stated preference literature, largely (although not always) independent of insights from behavioral welfare economics. There may be opportunities for validity improvements to stated preference studies via more formal and comprehensive integration of insights from these two areas. Are there generalizable insights to be gained from behavioral economics on how epistemically favorable conditions for welfare analysis can be better identified or encouraged in stated preference surveys? Is there additional behavioral science rigor that can be applied to the "art" of stated preference survey design?

Beyond the benefits of more comprehensive integration and hybridization of these two areas of work, there may be other, narrower and more concrete approaches that might be borrowed by stated preference researchers. As an illustrative example, one might extent stated preference research using insights from the modern preference expression literature, which evaluates the deliberative decision-making processes to identify normative preferences such as "deliberative mini-publics" (Courant, 2022; Farrell et al., 2019) or citizen juries where people are informed in depth about all details of the decision problem and only afterwards make their decisions. Approaches of this type have been used previously in the stated preference literature

(e.g., Álvarez-Farizo & Hanley, 2006), but not in ways that capitalize on insights from behavioral economics. It might also be possible to use self-reports, for example about whether decisions are regretted, whether a respondent attended to all of the presented information which was relevant to them, or whether respondents rejected or adjusted valuation scenario information, to identify whether stated preference data can be interpreted as normative and preference revealing (Cameron et al. 2011; Alemu et al. 2013; Lew and Whitehead 2020; Beshears et al., 2008). Ancillary or follow-up questions in stated preference surveys are often used for such purposes (Krupnick & Adamowicz, 2007).

3.3.3. *Bernheim-Rangel framework and stated preferences*

Stated preference researchers might also follow a modified two-step procedure in line with the formal Bernheim-Rangel framework. To identify the welfare relevant domain in the first step, stated preference researchers might identify context-dependent preferences by focusing on the characteristics and effects of the choice architecture described in subsection 1.4. of the Supplementary Material. Analysts then could identify mistakes within each of the context-dependent preferences and design surveys to avoid these mistakes (or ignore mistaken decisions for the welfare analysis). This, of course, requires an ability to identify and disentangle response and preference variations due to mistakes from those due to changes in the choice frame.

With this in mind, it should not be the aim of stated preference surveys to avoid frame-related inconsistencies across choices, as these inconsistencies can be due to constructed preferences. Surveys should only be designed to avoid/identify *mistaken* choices. To achieve this, the studies should be transparent, in that options and consequences are stated directly, and complex or opaque inferences should be avoided. The studies should also be designed so that people understand the underlying principles of the decision frame (for example, the provision rule that determines the relationship between their responses and the provision of the hypothetical good). At least in concept, approaches such as these are broadly consistent with (though not identical to) best practices in the stated preference literature. For example, best practices call for clear descriptions of the goods or services to be valued, along with the provision rule and related features of the choice architecture. They also recommend sensitivity analysis to evaluate the possible effects of potentially influential factors on welfare estimates (Johnston et al. 2017), as well as guidance to use a "weight of evidence" from many studies to evaluate validity and reliability (Bishop & Boyle, 2019).

Moreover, tools such as eye tracking can be used to identify whether people make use of all relevant information (e.g., Balcombe et al., 2017; Fraser et al., 2021; Uggeldahl et al., 2016; Van Loo et al., 2018), and at times it can be possible to test whether people make correct inferences or interpret questions correctly, e.g., in manipulation checks.[11] Process tracking can be used to assess attention, memory, forecasting, and other psychological processes. Approaches such as these can be used to identify welfare relevant stated preferences. Although some approaches of this type have been applied in the stated preference literature (e.g., attention and attentiveness checks; Sandorf, 2019), these applications are scattered across the literature and lack a cohesive foundation.

Once the welfare relevant choices have been identified, welfare consequences can be calculated across different possible choice frames. If there are inconsistent choices (due to the context-dependent character of the preferences), lower and upper bounds of WTP or WTA measures can be calculated. It can be explicitly stated that no welfare judgements can be made within these bounds using the tools available. Of course, there are multiple challenges and questions to be addressed when applying methods of this type to stated preferences used to inform decisions. For example, in the presence of context-dependent preferences, how many and which decision frames or contexts should be considered when establishing upper and lower welfare bounds? Moreover, are there bounds on what might be considered a valid frame or context for such an analysis?[12] Hence, while approaches such as the Bernheim-Rangel framework might offer a potential, rigorous foundation for addressing behavioral anomalies in stated preference analysis, developing the associated theory and methods would not be a trivial endeavor.

4. Conclusions

Stated preference methods have been criticized for their susceptibility to behavioral anomalies. This paper has shown that many of these can be explained and organized using behavioral economic theories and that there is ample evidence for similar behavioral patterns in settings beyond stated preferences. This observation suggests that stated preference "anomalies" are

---

[11] As an illustration, one might test whether subjects correctly understood information provided in the stated preference survey and whether the degree of misunderstanding influenced preference estimates (e.g., Needham et al. 2018).

[12] For example, one can imagine situations where unscrupulous researchers with a vested interest in producing either low or high welfare estimates might seek to develop survey instruments purposefully designed to elicit either zero or exceptionally high welfare estimates, by intentionally encouraging protest responses of various types. Should extreme values of this type be incorporated within upper or lower bounds for welfare analysis? If not, how should one determine a valid versus invalid frame for the analysis?

not, in fact, anomalous. Indeed, they are to be expected once stated preference studies are interpreted from the perspective of behavioral economics.

Having identified that stated preference anomalies are also present in the real world suggests that the distinction between stated and revealed preferences is not the most important distinction when evaluating whether elicited preferences should be used for welfare analysis. Situational factors can influence judgments and decisions in stated preference studies and the real world. As such, it is not always valid to claim that preferences revealed by real-world choices are more indicative of normative preferences than stated preferences elicited in incentive-compatible conditions. The influence of normatively-irrelevant situational factors may well be larger in the real world than in more controlled stated preference studies. In the real world, for example, market forces reward deception and the manipulation of preferences (Akerlof & Shiller, 2015), and revealed preferences are rarely elicited without influence of normatively irrelevant context factors. Rather than relying on revealed preferences (where possible) in lieu of stated preferences, we suggest that it is essential to come up with better criteria to decide which preference data should be used in welfare calculations, and under what circumstances.

Behavioral welfare economics suggests several ways to deal with behaviors that are not consistent with the standard neoclassical model of rational choice. The prospective feasibility of such approaches is supported by the general consistency of these behavioral welfare economic approaches with many current practices in the stated preference literature. We encourage future research to identify which of these approaches is the most applicable to stated preference research and to develop solid methodological foundations for applying these approaches in practice.

References

Akerlof, G. A., & Shiller, R. J. (2015). *Phishing for phools: The economics of manipulation and deception*. Princeton University Press.

Alemu, M. H., Mørkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and Resource Economics*, *54*, 333–359.

Allcott, H., & Taubinsky, D. (2015). Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, *105*, 2501–2538.

Álvarez-Farizo, B., & Hanley, N. (2006). Improving the Process of Valuing Non-Market Benefits: Combining Citizens' Juries with Choice Modelling. *Land Economics*, *82*, 465–478.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, *100*, 464–477.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*, 73–106.

Balcombe, K., Fraser, I., Williams, L., & McSorley, E. (2017). Examining the relationship between visual attention and stated preferences: A discrete choice experiment using eye-tracking. *Journal of Economic Behavior & Organization*, *144*, 238–257.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, *27*, 173–196.

Benjamin, D. J. (2019). Chapter 2—Errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 69–186). North-Holland.

Bernheim, B. D. (2016). The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics. *Journal of Benefit-Cost Analysis*, *7*, 12–68.

Bernheim, B. D. (2021). In defense of behavioral welfare economics. *Journal of Economic Methodology*, *28*, 385–400.

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

Bernheim, B. D., & Taubinsky, D. (2018). Chapter 5—Behavioral Public Economics. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 381–516). North-Holland.

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, *92*, 1787–1794.

Bishop, R. C., & Boyle, K. J. (2019). Reliability and validity in nonmarket valuation. *Environmental and Resource Economics*, *72*, 559–582.

Bishop, R. C., & Boyle, K. J. (2021). On Adding-Up as a Validity Criterion for Stated-Preference Studies. *Environmental and Resource Economics*, *80*, 587–601.

Börger, T. (2013). Keeping up appearances: Motivations for socially desirable responding in contingent valuation interviews. *Ecological Economics*, *87*, 155–165.

Börger, T., Maccagnan, A., White, M. P., Elliott, L. R., & Taylor, T. (2022). Was the trip worth it? Consistency between decision and experienced utility assessments of recreational nature visits. *American Journal of Agricultural Economics*.

Boyle, K. J., Bishop, R. C., & Welsh, M. P. (1985). Starting point bias in contingent valuation bidding games. *Land Economics*, *61*, 188–194.

Brown, T. C., & Gregory, R. (1999). Why the WTA–WTP disparity matters. *Ecological Economics*, *28*, 323–335.

Busse, M. R., Pope, D. G., Pope, J. C., & Silva-Risso, J. (2015). The psychological effect of weather on car purchases. *The Quarterly Journal of Economics*, *130*, 371–414.

Camerer, C. F. (2004). Prospect theory in the wild: Evidence from the field. *Advances in Behavioral Economics*, 148–161.

Cameron, T. A., & Englin, J. (1997). Respondent experience and contingent valuation of environmental goods. *Journal of Environmental Economics and Management*, *33*, 296–313.

Campbell, D., Hensher, D. A., & Scarpa, R. (2011). Non-attendance to attributes in environmental choice analysis: A latent class specification. *Journal of Environmental Planning and Management*, *54*, 1061–1076.

Carlsson, F. (2010). Design of Stated Preference Surveys: Is There More to Learn from Behavioral Economics? *Environmental and Resource Economics*, *46*, 167–177.

Carlsson, F., Gravert, C., Johansson-Stenman, O., & Kurz, V. (2021). The Use of Green Nudges as an Environmental Policy Instrument. *Review of Environmental Economics and Policy*, *15*, 216–237.

Carson, R. T. (1997). Contingent valuation surveys and tests of insensitivity to scope. In *Determining the value of non-marketed goods* (pp. 127–163). Springer.

Carson, R. T. (2012). Contingent valuation: A practical alternative when prices aren't available. *Journal of Economic Perspectives*, *26*, 27–42.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*, 181–210.

Champ, P. A., Moore, R., & Bishop, R. C. (2009). A comparison of approaches to mitigate hypothetical bias. *Agricultural and Resource Economics Review*, *38*, 166–180.

Chater, N. (2018). Mind Is Flat. In *Mind Is Flat*. Yale University Press.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review*, *99*, 1145–1177.

Conlin, M., O'Donoghue, T., & Vogelsang, T. J. (2007). Projection bias in catalog orders. *American Economic Review*, *97*, 1217–1249.

Courant, D. (2022). Institutionalizing deliberative mini-publics? Issues of legitimacy and power for randomly selected assemblies in political systems. *Critical Policy Studies*, *16*, 162–180.

Cummings, R. G., & Taylor, L. O. (1999). Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *The American Economic Review*, *89*, 649–665.

Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., … Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management*, *63*, 73–91.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, *47*, 315–372.

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, *127*, 1–56.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*, 123–143.

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press.

Dold, M. F. (2018). Back to Buchanan? Explorations of welfare and subjectivism in behavioral economics. *Journal of Economic Methodology*, *25*, 160–178.

Ellison, G., & Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, *77*, 427–452.

Farrell, D. M., Curato, N., Dryzek, J. S., Geissel, B., Grönlund, K., Marien, S., … Rose, J. (2019). Deliberative mini publics: Core design features. *The Centre for Deliberative Democracy & Global Governance Working Paper Series*, *2019*.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*, 817–868.

Fraser, I., Balcombe, K., Williams, L., & McSorley, E. (2021). Preference stability in discrete choice experiments. Some evidence using eye-tracking. *Journal of Behavioral and Experimental Economics*, *94*, 101753.

Friedrichsen, J., König, T., & Schmacker, R. (2018). Social image concerns and welfare take-up. *Journal of Public Economics*, *168*, 174–192.

Gabaix, X. (2019). Chapter 4—Behavioral inattention. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 261–343). North-Holland.

Gabaix, X., & Laibson, D. (2006). Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *The Quarterly Journal of Economics*, *121*, 505–540.

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, *74*, 1365–1384.

Gravert, C., & Kurz, V. (2021). Nudging à la carte: A field experiment on climate-friendly food choice. *Behavioural Public Policy*, *5*, 378–395.

Haab, T. C., Interis, M. G., Petrolia, D. R., & Whitehead, J. C. (2013). From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation. *Applied Economic Perspectives and Policy*, *35*, 593–612.

Hanley, N., Kriström, B., & Shogren, J. F. (2009). Coherent arbitrariness: On value uncertainty for environmental goods. *Land Economics*, *85*, 41–50.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge, UK: Cambridge University Press.

He, J., & Zhang, B. (2021). Current Air Pollution and Willingness to Pay for Better Air Quality: Revisiting the Temporal Reliability of the Contingent Valuation Method. *Environmental and Resource Economics*, *79*, 135–168.

Herriges, J. A., & Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management*, *30*, 112–131.

Ho, D. E., & Imai, K. (2008). Estimating causal effects of ballot order from a randomized natural experiment: The California alphabet lottery, 1978–2002. *Public Opinion Quarterly*, *72*, 216–240.

Horowitz, J. K., & McConnell, K. E. (2002). A review of WTA/WTP studies. *Journal of Environmental Economics and Management*, *44*, 426–447.

Infante, G., Lecouteux, G., & 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*, 1–25.

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*, 159–186.

Jacquemet, N., Joule, R.-V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*, 110–132.

Johansson, P.-O. (1992). Altruism in cost-benefit analysis. *Environmental and Resource Economics*, *2*, 605–613.

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 461.

Johnston, R. J. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. *Journal of Environmental Economics and Management*, *52*, 469–481.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., … Scarpa, R. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*, 319–405.

Johnston, R. J., Schultz, E. T., Segerson, K., Besedin, E. Y., & Ramachandran, M. (2013). Stated preferences for intermediate versus final ecosystem services: Disentangling willingness to pay for omitted outcomes. *Agricultural and Resource Economics Review*, *42*, 98–118.

Johnston, R. J., Schultz, E. T., Segerson, K., Besedin, E. Y., & Ramachandran, M. (2017). Biophysical causality and environmental preference elicitation: Evaluating the validity of welfare analysis over intermediate outcomes. *American Journal of Agricultural Economics*, *99*, 1–23.

Johnston, R. J., Weaver, T. F., Smith, L. A., & Swallow, S. K. (1995). Contingent valuation focus groups: Insights from ethnographic interview techniques. *Agricultural and Resource Economics Review*, *24*, 56–69.

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, *22*, 57–70.

Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues. *Journal of Risk and Uncertainty*, *19*, 203235.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*, 263–292.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341.

Kahneman, D., Wakker, P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, *112(2)*, 375–405.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, *68*, 94–101.

Kling, C. L., Phaneuf, D. J., & Zhao, J. (2012). From Exxon to BP: Has some number become better than no number? *Journal of Economic Perspectives*, *26*, 3–26.

Kremer, M., Rao, G., & Schilbach, F. (2019). Chapter 5—Behavioral development economics. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 345–458). North-Holland.

Krupnick, A., & Adamowicz, W. L. (Vic). (2007). Supporting Questions in Stated-Choice Studies. In B. J. Kanninen (Ed.), *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice* (pp. 43–65). Dordrecht: Springer Netherlands.

Lacetera, N., & Macis, M. (2010). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior & Organization*, *76*, 225–237.

LaRiviere, J., Czajkowski, M., Hanley, N., Aanesen, M., Falk-Petersen, J., & Tinch, D. (2014). The value of familiarity: Effects of knowledge and objective signals on willingness to pay for a public good. *Journal of Environmental Economics and Management*, *68*, 376–389.

Lew, D. K., & Whitehead, J. C. (2020). Attribute non-attendance as an information processing strategy in stated preference choice experiments: Origins, current practices, and future directions. *Marine Resource Economics*, *35*, 285–317.

Lichtenstein, S., & Slovic, P. (2006). *The construction of preference*. Cambridge University Press.

List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, *20*, 241–254.

Little, J., & Berrens, R. (2004). Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta- Analysis. *Economics Bulletin*, *3*, 1–13.

Lloyd-Smith, P., & Adamowicz, W. (2018). Can stated measures of willingness-to-accept be valid? Evidence from laboratory experiments. *Journal of Environmental Economics and Management*, *91*, 133–149.

Lusk, J. L., & Norwood, F. B. (2009). An inferred valuation method. *Land Economics*, *85*, 500–514.

McConnell, K. E. (1997). Does altruism undermine existence value? *Journal of Environmental Economics and Management*, *32*, 22–37.

McFadden, D. (1994). Contingent Valuation and Social Choice. *American Journal of Agricultural Economics*, 689–708.

Meier, A. N., Schmid, L., & Stutzer, A. (2019). Rain, emotions and voting for the status quo. *European Economic Review*, *119*, 434–451.

Miller, J. M., & Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly*, 291–330.

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, *30*, 313–325.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*, 263–280.

Needham, K., Czajkowski, M., Hanley, N., & LaRiviere, J. (2018). What is the causal impact of information and knowledge in stated preference studies? *Resource and Energy Economics*, *54*, 69–89.

Ojea, E., & Loureiro, M. L. (2007). Altruistic, egoistic and biospheric values in willingness to pay (WTP) for wildlife. *Ecological Economics*, *63*, 807–814.

Penn, J. M., & Hu, W. (2018). Understanding hypothetical bias: An enhanced meta-analysis. *American Journal of Agricultural Economics*, *100*, 1186–1206.

Poe, G. L. (2016). Behavioral anomalies in contingent values and actual choices. *Agricultural and Resource Economics Review*, *45*, 246–269.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, *36*, 11–46.

Sakaguchi, H., Stewart, N., Gathergood, J., Adams, P., Guttman-Kenney, B., Hayes, L., & Hunt, S. (2022). Default effects of credit card minimum payments. *Journal of Marketing Research*, 00222437211070589.

Sandorf, E. D. (2019). Did You Miss Something? Inattentive Respondents in Discrete Choice Experiments. *Environmental and Resource Economics*, *73*, 1197–1235.

Scarpa, R., Gilbride, T. J., Campbell, D., & Hensher, D. A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, *36*, 151–174.

Schkade, D. A., & Payne, J. W. (1994). How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation. *Journal of Environmental Economics and Management*, *26*, 88–109.

Sunstein, C. R. (2020a). Behavioral welfare economics. *Journal of Benefit-Cost Analysis*, *11*, 196–220.

Sunstein, C. R. (2020b). *Hayekian Behavioral Economics* (SSRN Scholarly Paper No. ID 3714750). Rochester, NY: Social Science Research Network.

Swait, J., & Adamowicz, W. (2001). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, *86*, 141–167.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*, 39–60.

Thaler, R. H. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*, *12*, 183–206.

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. WW Norton & Company.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Tunçel, T., & Hammitt, J. K. (2014). A new meta-analysis on the WTP/WTA disparity. *Journal of Environmental Economics and Management*, *68*, 175–187.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Uggeldahl, K., Jacobsen, C., Lundhede, T. H., & Olsen, S. B. (2016). Choice certainty in Discrete Choice Experiments: Will eye tracking provide useful measures? *Journal of Choice Modelling*, *20*, 35–48.

Van Loo, E. J., Nayga, R. M., Jr, Campbell, D., Seo, H.-S., & Verbeke, W. (2018). Using eye tracking to account for attribute non-attendance in choice experiments. *European Review of Agricultural Economics*, *45*, 333–365.

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*, 145–171.

Vossler, C. A., & Zawojska, E. (2020). Behavioral drivers or economic incentives? Toward a better understanding of elicitation effects in stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *7*, 279–303.

Whitehead, J. C. (2016). Plausible responsiveness to scope in contingent valuation. *Ecological Economics*, *128*, 17–22.

Supplementary Material

1. The three pillars of behavioral economics

Research in behavioral economics aims to improve the psychological realism of the traditional economic model. To do so, behavioral economic findings are frequently organized according to three systematic deviations from the standard economic model: non-standard preferences, non-standard beliefs, and non-standard decision-making (DellaVigna, 2009; Rabin, 1998). In the last decade, interactions between scholars at the interface of economics, psychology, and cognate disciplines have broadened the analysis and started to investigate how these three deviations interact with characteristics of the choice environment (also called the choice architecture) to influence behavior and welfare. Sometimes called "supposedly irrelevant factors", these characteristics (e.g., defaults, reminders, and labels) have been shown to have strong and systematic effects on behavior, although traditional economic models deem them irrelevant (Thaler, 2015).

Below, we present a necessarily incomplete list of key insights from the behavioral economics literature on the three types of deviations from the standard economic model and the influence of the choice architecture on these deviations. We focus on insights that are particularly salient for our organization of stated preference anomalies and the subsequent discussion how approaches from behavioral welfare economics might be applied to stated preference welfare analysis. More detailed summaries of the underlying behavioral science and applications to other areas of research are presented elsewhere (Benjamin, 2019a; Carlsson, Gravert, Johansson-Stenman, & Kurz, 2021; DellaVigna, 2009; Dhami, 2016; Kremer, Rao, & Schilbach, 2019).

1.1. Non-standard preferences

Research into non-standard preferences typically focuses on three domains: social preferences, risk preferences, and time preferences. *Social preferences* reflect the observation that people do not always act completely selfishly. Instead, they are concerned with social welfare and fairness (Camerer & Thaler, 1995), give to charity (DellaVigna, List, & Malmendier, 2012), and give tips although they do not plan to visit the venue again (Azar, 2004). Moreover, people compare themselves with others (Frank, 1985), reciprocate by rewarding others who are generous and punishing others who behave poorly (Fehr & Gächter, 2000), and act in line with

the social norms of social groups with whom people identify (Akerlof & Kranton, 2000). Altruism is also associated with pro-environmental behavior in everyday life (Lades, Laffan, & Weber, 2021).[1]

Within the domain of decision-making under risk, much of the literature focuses on *reference-dependent preferences*. Rather than evaluating outcomes in absolute terms, people evaluate outcomes as gains or losses relative to a reference point, for example, as described in prospect theory (Kahneman & Tversky, 1979). Compared to these reference points, people are often more sensitive to losses than to gains. Moreover, the subjective probabilities of gains and losses are not the same as the objective probabilities, as people overweight very small risks and underweight very high risks. Loss aversion can explain a commonly observed tendency to stick to the status quo, thus making it less likely, for example, to switch from grey energy to green energy (see also the endowment effect, Kahneman, Knetsch, & Thaler, 1991).

Regarding time preferences, behavioral economists have shown that many people do not discount the future exponentially as suggested in standard economics (Samuelson, 1937), but rather overweight the present as in models of present bias (Laibson, 1997) or have declining discount rates as in hyperbolic discounting (Ainslie, 1975). These models of time-inconsistent decision-making also reflect the possibility that some people might have self-control problems and deviate from prior plans. For example, people might plan to retrofit their homes to save energy but due to present bias endlessly procrastinate on making necessary arrangements (Lades, Clinch, & Kelly, 2021).

1.2. Non-standard beliefs

Non-standard beliefs are most relevant when people make decisions under risk or uncertainty. For example, when making such decisions people need to form beliefs about the likelihood of different outcomes. Standard theory assumes that fully rational decision-makers consider all available information and use Bayes' rule to update beliefs as new information becomes available. Behavioral economists, however, have identified situations wherein people form systematically incorrect beliefs about these likelihoods. Three broad domains capture much of the literature on non-standard beliefs: incorrect predictions about oneself in the future, the representativeness heuristic, and the tendency to obtain utility directly from beliefs.

---

[1] Some of these observed behaviors can be accommodated via adaptations of utility functions within traditional economic theory, such as the incorporation of different motives for altruism within neoclassical utility functions (McConnell, 1997).

People sometimes form systematically biased beliefs when *predicting their own future* preferences or mental states. For example, projection bias describes the tendency to project one's current state (e.g., hunger or fatigue) into the future (Loewenstein, O'Donoghue, & Rabin, 2003; Read & Van Leeuwen, 1998). Mispredictions of what outcomes will provide utility and well-being in the future are also common (Gilbert & Wilson, 2000). Moreover, some people lack the self-awareness to correctly predict their own future present bias and self-control problems (O'Donoghue & Rabin, 1999). This naiveté can prevent people from taking actions to avoid future self-control problems. Relatedly, many people are overconfident and systematically overestimate their actual abilities, performance, level of control, information, or chance of success.

The *representativeness heuristic* describes peoples' tendency to respond to questions about how likely outcomes are by answering the much simpler question about how representative these outcomes are (Kahneman & Tversky, 1972). Representativeness describes what we think the most typical example of a particular event or object is. The representativeness heuristic can be used to explain, for example, why some people believe that small random samples are as representative as large samples, why people think that when a coin flip comes up heads the next flip is more likely to come up tails, the hot hand fallacy, base-rate neglect, and the conjunction fallacy (Benjamin, 2019b).

Belief-based utility is the utility people derive from holding beliefs (Molnar & Loewenstein, 2021). This deviates from the standard economic views that beliefs are only an input to decision-making and suggests that people may also derive direct utility from, for example, others' beliefs about ourselves (whether others think that we are nice or intelligent), our beliefs about ourselves (e.g., self-images of being moral), and our beliefs about the world (whether we consider something to be fair or anticipation of future outcomes). Consequences of such beliefs include strategically ignoring certain types of information, interpreting information in a biased way (e.g., confirmation bias), and ideological conformity.

1.3. Non-standard decision-making

Non-standard decision-making describes systematic biases that influence our behavior even in situations of certainty. We differentiate between four main areas of non-standard decision-making: limited attention, limited cognitive effort, social influences, and emotions.

Simon (1955) highlighted that people have *limited cognitive abilities and limited attention* and that they often satisfice rather than maximize. Bounded rationality recognizes

that people have limited cognitive resources and that—given the complexity of the world—it is not possible to consider all available information and be fully attentive in all decisions. Due to limited attention and memory, people focus on and process only a subset of all the potentially available information (Chetty, Looney, & Kroft, 2009; Gabaix, 2019; Lacetera, Pope, & Sydnor, 2012).

Especially when choices are complex, as it is the case for many choices with environmental relevance, people have a preference to *avoid spending cognitive effort*. This can lead to inertia and the status quo bias (Kahneman et al., 1991; Madrian & Shea, 2001). When people feel that they are lacking a resource such as money or time, they may become occupied with thoughts related to having too little and these thoughts occupy cognitive resources and thus influence choices and behaviors. This preoccupation with *scarcity* can impede cognitive functions (Mullainathan & Shafir, 2013). Decisions can also be influenced by social pressure and *social influence*, i.e., pressure from their reference group (Akerlof, 1991).

Finally, *emotions* can influence behavior, but are rarely considered in the standard economic model of decision-making. For example, emotions can influence how much people tip in restaurants, hunger influences consumption decisions, and arousal influences the probability of engaging in morally questionable behavior (Ariely & Loewenstein, 2006; Loewenstein, 1996, 2000). Anticipatory emotions such as anticipated regret can also influence behavior (Zeelenberg, 1999).

1.4. Characteristics of the choice architecture

Characteristics of the choice architecture can influence behavior in any setting but can be particularly influential when people deviate from the standard economic model in one or more ways described above (Thaler & Sunstein, 2008). Many public and private institutions have begun to apply behavioral scientific insights and have integrated a broad set of behavioral and psychological influences into their policy design (OECD, 2017). Applied behavioral work has identified many contextual factors in field experiments and demonstrated that people can be "nudged" to change behavior. This literature suggests, for example, a prominent role for default effects on human behavior (Jachimowicz, Duncan, Weber, & Johnson, 2019), and noticeable, if more modest, effects of simplification, personalization, implementation intention prompts, reminders, and social norm comparisons (Carlsson et al., 2021; DellaVigna & Linos, 2020). Below, we focus on four groups of context factors: defaults, simplification and salience, psychologically informed messaging, and sludge.

Defaults are pre-selected options. Especially when people do not have a strong preference for one option over the other, defaults can increase the likelihood of the pre-selected option being chosen. Changing defaults is arguably the strongest intervention behavioral scientists have implemented (Jachimowicz et al., 2019). Defaults can have strong effects in domains such as organ donation, pension participation, energy provider choice, and printing, among others (Beshears, Choi, Laibson, & Madrian, 2009; Egebark & Ekström, 2016; Johnson & Goldstein, 2012; Kaiser, Bernauer, Sunstein, & Reisch, 2020; Pichert & Katsikopoulos, 2008). Kaiser et al. (2020) identify seven different reasons for the effectiveness of defaults, including status quo bias and inertia, satisficing behavior, and the endowment effect.

Simplification describes the process of communicating similar information in simpler terms and one can increase s*alience* by orienting people's attention to particular dimensions of decision. These actions can influence behavior because humans have limited attention and may not invest the time and effort required to consider all information within complex decisions. For example, Sunstein used "plate, not pyramid" as a mantra when working in the US Obama administration, referring to the change from a complex food pyramid to a simple plate to communicate healthy eating options to the public (Sunstein, 2013). Energy performance certificates and building energy ratings have been simplified worldwide to make it easier for buyers and homeowners to understand and compare energy-efficiency investments. Eco-labeling and fuel energy labels are ways to orient consumers' attention to the environmental relevance of purchases while saving money at the same time (Allcott & Sunstein, 2015). The order in which dinner options are presented on a restaurant menu can also influence food choices (Gravert & Kurz, 2021).

Applied behavioral scientists have identified a range of *psychologically informed messages* that influence environmental behavior. These messages are provided in letters, emails, phone messages, websites, etc. For example, informing people about what others do (i.e., social norm comparisons) in home energy reports and messages about towel reuse in hotel rooms can change behavior (Allcott, 2011; Goldstein, Cialdini, & Griskevicius, 2008). Other messages that can affect environmentally relevant behavior to varying degrees include moral appeals (Egebark & Ekström, 2016; Ito, Ida, & Tanaka, 2018), warnings and reminders (e.g., messages to reduce heating), and prompts (Ghesla, Grieder, Schmitz, & Stadelmann, 2020; Gollwitzer, 1999; Harding & Hsiaw, 2014).

Finally, behavioral scientists use the term *sludge* to describe excessive or unjustified frictions that make it harder for people to do what they wish, similar to transaction costs in standard economic models (Sunstein, 2020b; Thaler, 2018). Sludging is the "dark cousin" of

nudging, or nudging for evil (Thaler, 2018). Sludge makes it more difficult for people to navigate through their everyday lives (Sunstein, 2019). Sludge is present when, for example, paperwork is unnecessarily complicated, when forms need to be filled although the relevant information is already provided, when a website's design makes it difficult to identify the final price of a product, when the fine print it long and confusing, when defaults are set unfavorably, when messages induce psychological costs (Shahab & Lades, 2021; Soman, 2020; Sunstein, 2020).

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, *82*, 463–496.

Akerlof, G. A. (1991). Procrastination and Obedience. *The American Economic Review*, *81*, 1–19.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity. *Quarterly Journal of Economics*, *115*, 715–753.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*, 1082–1095.

Allcott, H., & Sunstein, C. R. (2015). Regulating Internalities. *Journal of Policy Analysis and Management*, *34*, 698–705.

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*, 87–98.

Azar, O. H. (2004). What sustains social norms and how they evolve?: The case of tipping. *Journal of Economic Behavior & Organization*, *54*, 49–64.

Benjamin, D. J. (2019a). Chapter 2—Errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 69–186). North-Holland.

Benjamin, D. J. (2019b). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, *2*, 69–186.

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2009). The importance of default options for retirement saving outcomes: Evidence from the United States. In *Social security policy in a changing environment* (pp. 167–195). University of Chicago Press.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, *9*, 209–219.

Carlsson, F., Gravert, C., Johansson-Stenman, O., & Kurz, V. (2021). The Use of Green Nudges as an Environmental Policy Instrument. *Review of Environmental Economics and Policy*, *15*, 216–237.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review*, *99*, 1145–1177.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, *47*, 315–372.

DellaVigna, S., & Linos, E. (2020). *Rcts to scale: Comprehensive evidence from two nudge units*. National Bureau of Economic Research.

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, *127*, 1–56.

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press.

Egebark, J., & Ekström, M. (2016). Can indifference make the world greener? *Journal of Environmental Economics and Management*, *76*, 1–13.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*, 159–181.

Frank, R. H. (1985). The demand for unobservable and other nonpositional goods. *American Economic Review*, *75*, 101–116.

Gabaix, X. (2019). Chapter 4—Behavioral inattention. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 261–343). North-Holland.

Ghesla, C., Grieder, M., Schmitz, J., & Stadelmann, M. (2020). Pro-environmental incentives and loss aversion: A field experiment on electricity saving behavior. *Energy Policy*, *137*, 111131.

Gilbert, D. T., & Wilson, T. D. (2000). Miswanting: Some problems in the forecasting of future affective states. *Feeling and Thinking: The Role of Affect in Social Cognition*, 178–197.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*, 472–482.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*, 493.

Gravert, C., & Kurz, V. (2021). Nudging à la carte: A field experiment on climate-friendly food choice. *Behavioural Public Policy*, *5*, 378–395.

Harding, M., & Hsiaw, A. (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization*, *107*, 209–227.

Ito, K., Ida, T., & Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, *10*, 240–267.

Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, *3*, 159–186.

Johnson, E. J., & Goldstein, D. G. (2012). *Decisions by Default* (SSRN Scholarly Paper No. ID 2173387). Rochester, NY: Social Science Research Network.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, 193–206.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*, 263–292.

Kaiser, M., Bernauer, M., Sunstein, C. R., & Reisch, L. A. (2020). The power of green defaults: The impact of regional variation of opt-out tariffs on green energy demand in Germany. *Ecological Economics*, *174*, 106685.

Kremer, M., Rao, G., & Schilbach, F. (2019). Chapter 5—Behavioral development economics. In B. D. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1* (pp. 345–458). North-Holland.

Lacetera, N., Pope, D. G., & Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, *102*, 2206–2236.

Lades, L. K., Clinch, J. P., & Kelly, J. A. (2021). Maybe tomorrow: How burdens and biases impede energy-efficiency investments. *Energy Research & Social Science*, *78*, 102154.

Lades, L. K., Laffan, K., & Weber, T. O. (2021). Do economic preferences predict pro-environmental behaviour? *Ecological Economics*, *183*, 106977.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, *112*, 443–477.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*, 272–292.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, *90*, 426–432.

Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection Bias In Predicting Future Utility. *Quarterly Journal of Economics*, *118*, 1209–1248.

Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, *116*, 1149–1187.

McConnell, K. E. (1997). Does altruism undermine existence value? *Journal of Environmental Economics and Management*, *32*, 22–37.

Molnar, A., & Loewenstein, G. (2021). Thoughts and players: An Introduction to old and new economic perspectives on beliefs. *The Science of Beliefs: A Multidisciplinary Approach (Provisional Title, to Be Published in October 2021). Cambridge University Press. Edited by Julien Musolino, Joseph Sommer, and Pernille Hemmer*.

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

O'Donoghue, T., & Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, *89*, 103–124.

OECD. (2017). Behavioural Insights and Public Policy—Lessons from Around the World—En—OECD. Retrieved October 18, 2018, from http://www.oecd.org/gov/regulatory-policy/behavioural-insights-and-public-policy-9789264270480-en.htm

Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, *28*, 63–73.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, *36*, 11–46.

Read, D., & Van Leeuwen, B. (1998). Predicting Hunger: The Effects of Appetite and Delay on Choice. *Organizational Behavior and Human Decision Processes*, *76*, 189–205.

Samuelson, P. A. (1937). A note on measurement of utility. *Review of Economic Studies*, *4*, 155–161.

Shahab, S., & Lades, L. K. (2021). Sludge and transaction costs. *Behavioural Public Policy*, 1–22.

Soman, D. (2020). Sludge: A Very Short Introduction. *BEAR.[Online][Date Accessed: 18/09/2020]: Https://Www. Rotman. Utoronto. ca/-/Media/Files/Programs-and-Areas/BEAR/White-Papers/BEARxBIOrg-Sludge-Introduction. Pdf*.

Sunstein, C. R. (2013). *Simpler: The future of government*. Simon and Schuster.

Sunstein, C. R. (2019). *On Freedom*. Princeton University Press.

Sunstein, C. R. (2020). Sludge audits. *Behavioural Public Policy*, 1–20.

Thaler, R. H. (2015). *Misbehaving: The Making of Behavioral Economics*. WW Norton & Company.

Thaler, R. H. (2018). Nudge, not sludge. *Science*, *361*, 431–431.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, *12*, 93–106.