Consumption of Values*

Itzhak Gilboa[†], Stefania Minardi,[‡] and Fan Wang[§]

May 20, 2021

Abstract

Consumption decisions are partly influenced by values and ideologies. Consumers care about global warming, child labor, fair trade, etc. Incorporating values into the consumer's utility function will often violate monotonicity, if consumption hurts values in a way that isn't offset by hedonic benefits. We distinguish between intrinsic and instrumental values, and argue that the former tend to introduce discontinuities near zero. For example, a vegetarian's preferences would be discontinuous near zero amount of animal meat. We axiomatize a utility representation that captures such preferences and discuss the measurability of the degree to which consumers care about such values.

*We thank Rossella Argenziano, Roy Fabian, Drew Fudenberg, Alma Gilboa-Schechtman, Brian Hill, Edi Karni, R. Vijay Krishna, Eric Maskin, John Quah, Ariel Rubinstein, David Schmeidler, Ran Spiegler, Tomasz Strzalecki, as well as participants of seminars at BIU, FSU, Harvard-MIT, JHU, NYU, and Penn for comments and references. We gratefully acknowledge support from the Investissements d'Avenir ANR -11- IDEX-0003 / Labex ECODEC No. ANR - 11-LABX-0047 as well as ISF Grant 1077/17, the AXA Chair for Decision Sciences at HEC, the Foerder Institute at Tel-Aviv University and the Sapir Center for Economic Developmemt (Gilboa).

[†]Tel-Aviv University and HEC, Paris. tzachigilboa@gmail.com – Corresponding Author

[‡]HEC, Paris. minardi@hec.fr

[§]ESSEC, Singapore. b00786349@essec.edu

1 Introduction

1.1 Motivation

In November 2015 Volkswagen sales in the US were about 25% lower than the year before. This dramatic drop followed a notice by the United States Environmental Protection Agency about the car manufacturer's violation of the Clean Air Act. It stands to reason that consumers were reacting to the facts that Volkswagen was selling cars that polluted the air beyond the allowed limits, and was also deceitful about it. Importantly, the information revealed about the cars and about the company's conduct had little to do with the very experience of consumption or even with its long term effects on the consumers themselves. Rather, it appears that consumers felt that two values were compromised by the firm's conduct: minimizing pollution and being honest. Consumers might have been angry at Volkswagen for its choices. Alternatively, they might have just decided not to be part of a deal that does not respect these values. Many consumers who decided not to buy a Volkswagen may have had a combination of the emotional reaction and the moral choice. In any event, this is a consumption choice that was partly determined by values.

Along similar lines, Nike has been struggling with information and rumors about its production practices for decades. In the 1990s it was reported that the company had been using sweatshop and child labor. Nike made a major effort to clean up its image, in an attempt to avoid the negative impact on sales. Again, whether or not child labor is involved in the production process does not affect the quality of the shoes or the experience of running with them. Rather, it had to do with what consumer perceived as the right choice: using child labor is considered immoral.[1]

[1]Nike argued that it had no control over the practices employed by its sub-contractors. We make no claim about Nike's actual conduct in this case, nor about Volkswagen's in the previous one. We only point out that consumers seem to care about values, and perceived disrespect for values can affect consumption choices.

These are but two examples in which consumers care not only about the product they get for their money, but also about values, and, in particular, about potential conflict between their consumption and values they hold. Some people prefer to consume only vegetarian or vegan products, while others would only consume Halal or Kosher food. Many consumption decisions are affected by the degree to which the production and/or the consumption processes hurt wildlife and endangered species, the globe and sustainability of life on it, or help underprivileged populations, promote equality, and so forth. These considerations are seldom the only ones consumers care about, and not all consumers care about them to the same degree. Yet, they can have non-negligible effects on consumer choices. For example, De Pelsmacker, Driesen, and Rayp (2005) found that consumers expressed a higher willingness to pay for coffee that was labeled "Fair Trade", while Hainmueller, Hiscox, and Sequeira (2015) showed that the label increased market share in a field experiment. Such ethical concerns affect firms' decisions. Indeed, the concept of Corporate Social Responsibility (CSR) might be partly a response to consumers' demand for values (see Garigga and Mele, 2004, for a survey of CSR theories).

1.2 Goal

By and large, economic theory tends to ignore value considerations. Classical economic textbooks (Varian, 1978, Kreps, 1990, Mas-Colell-Whinston-Green, 1995) start by conceptualizing a consumer's utility as a function of her own bundle. They proceed to deal with externalities, where one's consumption choices directly affect another's utility. The standard examples of externalities deal with the physical impact of goods consumed, as in the cases of contributions to public goods, pollution, etc. It is rarely the case that the values that consumption supports or hurts factor into the utility function. As described in the next subsection, values and meaning have been discussed extensively in a variety of fields, including, but not limited to, applied eco-

nomics and marketing. However, very little seems to have been done in terms of incorporating values into microeconomic theory, in terms of a formal, axiomatically-based model of consumer choice where consumers derive utility not only from material bundles, but also from values.

Economists might wonder whether a formal model of values is needed at all. If we only wish to understand economic behavior, one may argue that the values economic agents have, the degree to which they care about these values, and their willingness to trade off material convenience for preservation of these values are all implicitly captured by the utility function. After all, the utility function is behaviorally defined; if agents do indeed care about values in ways that affect their economic choices, a utility function that represents these choices would automatically incorporate the underlying values. It would thus seem that no new theory of values is needed in order to understand economic behavior.

We find this conclusion premature. First, it is important to study the way that hedonic and value utility are combined into the overall utility function. In particular, we would like to know when the utility can be assumed additively separable, whether its components are continuous etc. Analyzing the structure of the utility function adds to our theoretical understanding, and may also be useful in empirical estimation. Second, economic behavior that is value-driven can change as a result of information that has nothing to do with the product quality or the experience of its consumption per se. For example, information about flight gas emission can change consumption patterns, make some agents travel less or use trains rather than airplanes. Consumption data, given different informational states, can provide more reliable estimates of the importance of values than, say, questionnaires. But this requires that we open the "black box" of the utility function and study the way it changes in response to value-relevant information.[2]

[2]In principle, one may treat this information as part of a Bayesian model, and study consumer preferences before and after the arrival of information about the material goods. Such a model, however, would be hardly intuitive, and would require considerably more

The general framework we have in mind employs an additive utility function. We assume that the consumer is given information about goods' features, and maximizes a utility function given that information. The information might be represented by a vector d, where a number d_i describes the degree that product i hurts or supports a value. Given such a d, the consumer will be characterized by a preference order $\succsim_d$ represented by a function u_d on consumption bundles. For simplicity we focus on additive models, where this function takes the form

$$u_d(x) = u(x) + v(d, x)$$

such that, for each bundle x, $u(x)$ is the hedonic utility derived from the material goods in x, and, for each information state d, $v(d, x)$ is the (dis)utility that results from the effects the bundle x has on the consumer's values. These effects might have to do with production and/or with consumption of the goods involved. For example, in the case of child labor, the problem lies in the production process; by contrast, in the case of Kosher food, it is the consumption of the good which generates the negative effect on values. In other cases, such as vegetarianism, consumers may have negative reactions both to the production and to the consumption processes: most vegetarians oppose the killing of animals, but also the consumption of meat of animals that "naturally" died. In our model we do not attempt to distinguish among these, and, relatedly, also not between negative emotions that are invoked by the act of purchasing and by the act of consuming a good. We implicitly assume that the consumer is rational enough not to buy goods that she doesn't consume, and the information state d should describe the value-relevant information on the goods incorporating both production and consumption effects. Similarly, the function $v(d, x)$ should capture both the negative effects of purchase or ownership of the good and those of consuming

data to estimate the utility function as well as beliefs about the degree to which goods affect values.

it.[3]

Observe that in our conceptualization, preferences over bundles x are assumed observable *given* information states d, but we do not assume preference over these states, nor preferences over bundles across different information states d. Moreover, in some cases it will be unnatural to assume preferences over all bundles given any state d. For example, if d is a vector that indicates which goods contain animal meat and which are vegetarian, it might be artificial to assume that we can observe preferences under the assumption that beef is vegetarian. By contrast, in the example of the "Fair Trade" label, we may observe choices of the very same products with or without the label.[4]

Incorporating values into the consumer's utility function calls into question two of the basic properties of consumer preferences: monotonicity and continuity. Monotonicity might be violated because, in many examples, $v(d, x)$ will be decreasing in x. Consider, for example, a vegetarian consumer who prefers not to consume and not to own meat, even at zero cost. Increasing the amount of meat in her bundle will lower her utility. Similarly, a consumer who feels bad about CO_2 emissions may feel worse should her flights increase the level of global emission, and she may well reach a region in the bundle space where her preferences decrease in the quantity of flights.[5] The standard rationale for monotonicity is free disposal: a consumer need not physically consume products that she legally owns. But in the presence

[3]For further distinctions and extensions of the model, see subsection 5.2.

[4]There are values that introduce mixed cases. Consider, again, the example of Kosher food. There are products that the consumer would always wish to avoid, such as pork. Asking the consumer to report her preferences under the assumption that pork were Kosher would be rather fanciful. But there are products that may or may not be Kosher, depending on external information. For example, if a product is sold by a store that is owned by Jews and that opens on Saturdays, the product is non-Kosher. If the same store is known to keep Kosher (and to observe Saturday), the product may be Kosher as well. Our main interest is, however, in the former case, which is more challenging in terms of the information one may assume available.

[5]Along similar lines, value-dependent preferences may also violate local non-satiation. For example, if all goods involve some environmental damage, the consumer may reach a maximum of the utility function which is in the interior of her budget set.

of values free disposal no longer holds. A person might feel guilty about the degree to which the bundle she owns hurts certain causes. Because there is no free disposal of emotions, preferences need not be monotone.[6]

We distinguish between two types of values: *intrinsic* and *instrumental*. The former are ends in themselves, while the latter are proxies for other, "ultimate", or "pure" values. For example, avoiding child labor is probably an intrinsic value for most consumers: people typically do not frown upon child labor only or mostly because it has negative long-term effects; rather, it just feels wrong. By contrast, minimizing carbon dioxide emission is hardly a value in its own right. Having this or that gas in the atmosphere is, in itself, morally neutral. Minimizing emissions is a value only as a proxy for the underlying value of preserving the planet and, in turn, for the (ultimate, intrinsic) value of taking future generations into account in our consumption decisions.

This paper suggests to distinguish between intrinsic and instrumental values along the lines of continuity: an intrinsic value is compromised as soon as it is violated to *some* positive degree, no matter how small. Buying a product that is known to have been produced employing child labor feels wrong, whether the amount of labor involved was large or small. A vegetarian consumer would wonder whether a bundle is vegetarian, and if it isn't, the amount of animal meat in it doesn't seem to matter that much. Similarly, an observant Jew who only eats Kosher food categorizes bundles in a dichotomous way. By contrast, instrumental values tend to be judged in a continuous way. One may wish to avoid consumption that generates greenhouse gas emission, but if the amount of gas emission is negligible, so will the emotional impact of consumption be.[7]

[6]One may argue that a value is, by definition, something for which the agent is willing to give up hedonic well-being. This, however, does not necessarily imply violation of monotonicity, because an increase in consumption quantities x may lead to an increase in hedonic well-being ($u(x)$) that is enough to offset the negative impact this consumption has on values ($v(d,x)$).

[7]The distinction between intrinsic and instrumental values is subjective. For instance,

We therefore conceptualize intrinsic values as related to discontinuities of the function v near zero; for example, if $x_i = 0$ for all goods i that contain animal meat, $v(d, x)$ might be 0, and if $x_i > 0$ for some of them – $v(d, x)$ assumes a negative value, bounded away from zero. The source of this discontinuity is the mental act of assigning *meaning* to consumption. Whereas our bodily perceptions tend to be continuous in quantities, the meaning that we attach to physical bundles is not. In the case of intrinsic values, the goods themselves are the carriers of meaning, and thus we expect discontinuities (in v and therefore in u_d) to arise. But in the case of instrumental values, the goods are only proxies; they affect the truly meaningful values via some mechanism, which may be physical, biological, sociological etc. Since these mediating mechanisms tend to be continuous, we expect the preferences of a rational, well-informed consumer to be continuous as well.

The simplest model of intrinsic values will therefore involve a function $v(d, x)$ that may assume only two values. Such dichotomous values will be dubbed *principles*. If the consumer has but a single principle, the relevant information d is simply a vector of binary components, where $d_i \in \{0, 1\}$ indicates whether good i violates the principle or not. For example, $d_i = 0$ indicates that good i is vegetarian, whereas $d_i = 1$ – that it isn't.[8] A consumption bundle x is then evaluated by[9]

$$u_d(x) = u(x) - \gamma \mathbf{1}_{\{d \cdot x > 0\}} \tag{1}$$

where $\gamma \geq 0$ measures the degree to which the consumer cares about the principle.

---

some consumers may avoid the use of plastic bags as a matter of principle. Others may feel that these consumers are "too religious" about plastic bags. Thus, we find that (dis)continuity at zero is a reasonable test to tell apart intrinsic from instrumental values, whether, for a given consumer, the value is intrinsic or instrumental.

[8] Obviously, this is a simplified model. A vegetarian might still distinguish between beef and fish, and prefer eating seafood to mammals.

[9] Fehr and Schmidt (1999) also use, and Karni and Safra (2002) axiomatize utility functions that are additively separable between hedonic utility and a component that represents values of equality or fairness.

The next subsection provides an example of such preferences. It is mostly supposed to illustrate the way that γ can be elicited from observed choice. Our main result is the axiomatization of preferences as in (1), provided in Section 2. We assume a given information state d and provide conditions on a binary relation that can be represented by u_d as above (where u is continuous but $\gamma > 0$ introduces discontinuity). Section 3 offers some extensions of the model, including a simple example of an instrumental value. A survey of related literature is provided in 4. Section 5 concludes with a general discussion.

1.3 Example

Consider a consumer problem with two goods: vegetables and meat. Let x_1 and x_2 denote their quantities, and assume first a utility function

$$u(x_1, x_2) = \alpha \log(x_1 + x_2) + x_2$$

with $\alpha > 0$. The first component, $\alpha \log(x_1 + x_2)$, captures the satisfaction of hunger, for which the two goods have the same impact. The second component is designed to capture some of the reasons for which people like meat: the nutritional content, the taste, or evolution that shaped the latter to match the former. The consumer faces a standard budget constraint

$$p_1 x_1 + p_2 x_2 \leq I$$

and we assume that $p_1 < p_2$. It can be verified that the optimal solution is:

(i) For low income, $I \leq \alpha(p_2 - p_1)$, the solution is $\left(\frac{I}{p_1}, 0\right)$;

(ii) For high income, if $I \geq \alpha \frac{p_2}{p_1}(p_2 - p_1)$, it is $\left(0, \frac{I}{p_2}\right)$;

(iii) In between, if $\alpha(p_2 - p_1) < I < \alpha \frac{p_2}{p_1}(p_2 - p_1)$, it is given by

$$\begin{aligned} x_1 &= -\frac{I}{p_2 - p_1} + \alpha \frac{p_2}{p_1} \\ x_2 &= \frac{I}{p_2 - p_1} - \alpha \end{aligned} \tag{2}$$

Let us now introduce a principle into the picture. Suppose that the consumer cares about animals and feels better thinking that no animal had to be killed for her meal. Specifically, we assume that the consumer maximizes the function

$$u_d(x_1, x_2) = \alpha \log(x_1 + x_2) + x_2 - \gamma \mathbf{1}_{\{x_2>0\}} \tag{3}$$

in which a penalty $\gamma \geq 0$ is deducted from the utility of a bundle (x_1, x_2) if and only if x_2 is consumed at a positive level. Let us distinguish among three cases:

(i) $I \leq \alpha(p_2 - p_1)$ so that $\left(\frac{I}{p_1}, 0\right)$ is a maximizer of u. It follows that it is the unique maximizer of u_d for any $\gamma \geq 0$: a person who anyway chose (or could have chosen) not to consume meat without being vegetarian will certainly not consume meat if she became vegetarian.

(ii) $\alpha(p_2 - p_1) < I < \alpha\frac{p_2}{p_1}(p_2 - p_1)$ and the maximizer of u is defined by (2). It is also optimal for u_d if and only if

$$\alpha \log\left(\alpha\frac{p_2 - p_1}{I}\right) + \frac{I}{p_2 - p_1} - \alpha \geq \gamma \tag{4}$$

Notice that, for positive γ, there will be ranges of income $I > \alpha(p_2 - p_1)$ for which the inequality will not hold. In other words, a consumer who cares about the vegetarian principle to degree $\gamma > 0$ would start consuming meat later than would a consumer who doesn't care about this value. As long as meat is more expensive than are vegetables, both would consume only vegetables for very low income values (as in (i) above). However, when they get richer, the consumer who cares about the value would refrain from consuming meat up to a higher income level than would the consumer who doesn't.

Notice that the LHS of (4) is unbounded in I. This means that, for any value of γ, there will be a high enough income level for which the inequality would hold. However, whether (2) is the optimal solution depends also on I not being too high as to leave the range in which (2) applies.

(iii) Finally, if $\alpha\frac{p_2}{p_1}(p_2 - p_1) \leq I$, $\left(0, \frac{I}{p_2}\right)$ is a maximizer of u. It is also a maximizer of u_d if and only if

$$\alpha \log\left(\frac{p_1}{p_2}\right) + \frac{I}{p_2} \geq \gamma \tag{5}$$

Observe that this inequality holds in this income range for $\gamma = 0$: the range is defined by $\frac{I}{p_2} \geq \alpha\frac{p_2 - p_1}{p_1} = \alpha\left(\frac{p_2}{p_1} - 1\right)$ and, for $\frac{p_2}{p_1} > 1$, we have $\frac{p_2}{p_1} - 1 > \log\left(\frac{p_2}{p_1}\right)$, so that $\alpha\frac{p_2 - p_1}{p_1} > -\alpha\log\left(\frac{p_1}{p_2}\right)$ and hence (5) holds for $\gamma = 0$.

Importantly, this analysis allows for measurement of γ.

In this example we found that, for any $\gamma > 0$, there is a high enough income for which x_2 is consumed at a positive level. Because of the additive structure of the model and the fact that the hedonic utility u was unbounded, a high enough income made vegetarianism "too expensive" in terms of its opportunity cost. This might appear somewhat cynical, describing a consumer who is always willing to compromise her principle for a sufficiently high hedonic benefit. By contrast, if u were bounded and we had

$$\gamma > \sup(u) - \inf(u)$$

we would find that the consumer would never give up her principle. Such a consumer would starve rather than have non-vegetarian food. Finally, if u is bounded from above but not from below, we can model a consumer who would eat meat in order to survive but not otherwise.[10] We now turn to describe and axiomatize the general model.

[10] Clearly, if $\gamma > \sup(u) - \inf(u)$ the parameter γ cannot be identified, as all such values of γ lead to the same observed choice. A one-sided infinite range of u, by contrast, would still allow for the identification of γ.

2 Axiomatization

2.1 Set-up

The alternatives are consumption bundles in X, which is a closed and convex subset of $\mathbb{R}^n_+$. For each good $i \leq n$ there is an indicator $d_i \in \{0,1\}$ denoting whether the good violates the principle. That is, $d_i = 1$ implies that the good is inconsistent with the principle (say, contains meat), and $d_i = 0$ – that it doesn't (purely vegetarian). The consumer is aware of the vector $d \in \{0,1\}^n$, where we assume that producers should and do truthfully disclose the ingredients of their products.

We wish to axiomatize the model in which, given d, the consumer maximizes

$$u_d(x) = u(x) - \gamma \mathbf{1}_{\{d \cdot x > 0\}} \tag{6}$$

where $d \cdot x$ is the inner product of the two vectors, so that $d \cdot x > 0$ if and only if there exists a product i that violates the principle ($d_i = 1$) and that is consumed at a positive quantity in x.

In this section we assume that the vector d is known and kept fixed. That is, the consumer is provided with information about the goods that are and are not vegetarian, and we implicitly assume that this information is trusted. We keep the information fixed, and can therefore suppress d from the notation, assuming that a binary relation $\succsim_d = \succsim \subset X \times X$ is observable. The information contained in the vector d is summarized by the answer to the question, is $d \cdot x > 0$? We thus define

$$X^0 = \{\, x \in X \mid d \cdot x = 0 \,\}$$

that is, all consumption bundles that do not use any positive amount of the "forbidden" goods, while

$$X^1 = X \backslash X^0 = \{\, x \in X \mid d \cdot x > 0 \,\}$$

contains the other bundles. Observe that X^0 is closed and convex and X^1 is convex.

Before moving on, we introduce some notation. The term "a sequence $(x_n)_{n\geq 1} \to_{n\to\infty} x$" will refer to a sequence $(x_n)_{n\geq 1}$ such that $x_n \in X$ for all n, and $x_n \to_{n\to\infty} x$ in the standard topology, where $x \in X$. When no ambiguity is involved, we will omit the index notation "$n \to \infty$" as well as the subscript "$n \geq 1$". We will use the notation "a sequence $(x_n) \subset A$" for "a sequence $(x_n)_{n\geq 1}$ such that $\{x_n\}_{n\geq 1} \subset A$".

Conditions that involve an unspecified index such as $x_n \succsim y_n$ are understood to use a universal quantifier ("for all $n \geq 1$"). Finally, when no confusion is likely to arise we will also omit the parentheses and use $x_n \to x$ rather than $(x_n) \to x$.

2.2 Axioms

We impose the following axioms on $\succsim$. We start with the standard assumption positing that choice behavior is described by a complete preorder.

A1. Weak Order: $\succsim$ is complete and transitive on X.

The next axioms will make use of the following key notion:

Definition 1 *Two sequences $x_n \to x$ and $y_n \to y$ are* comparable *if*

(A) there exist $i, j \in \{0, 1\}$ such that $(x_n) \subset X^i, x \in X^i$ and $(y_n) \subset X^j, y \in X^j$

or

(B) there exist $i, j \in \{0, 1\}$ such that $(x_n), (y_n) \subset X^i$ and $x, y \in X^j$.

Clearly, if all of the elements of $(x_n), (y_n)$, as well as the limit point of each are in the same subspace – X^0 or X^1 – the sequences are comparable.[11] However, two sequences $x_n \to x$ and $y_n \to y$ are comparable also in two other cases: first, (A) if (x_n) as well as *its* limit x are all in one subspace, while (y_n) with *its* limit, y, are all in another. And, second, (B) if the elements of both sequences belong to X^1 and the limits of both belong to X^0. (In

[11] Here and in the sequel we use the terms "space" and "subspace" in the topological sense.

principle, the opposite is also allowed by the definition, but X^0 is closed, so we cannot have a sequence in it converging to a point in X^1.) Basically, comparability rules out cases in which the transition to the limit makes only one sequence cross the boundary between the subspaces, leaving X^1 and reaching X^0. If this occurs, then the information we gather from preferences along the sequences is not very useful for making inferences about the limits: one sequence changes in a way that is discontinuous, and the other one doesn't. By contrast, if the two sequences are comparable because none of them crosses the boundary between the two subspaces, then there is no reason for any violation of continuity. And, importantly, if both do cross the boundary, we still expect preference information along the sequences (where both (x_n) and (y_n) are in one subspace, which can only be X^1 in this case) to carry over to the limits (even though these are located in another subspace).

We can now state our continuity axiom:

A2. Weak Preference Continuity: For all comparable sequences $x_n \to x$ and $y_n \to y$, if $x_n \succsim y_n$ then $x \succsim y$.

Observe that, without the comparability condition, A2 would be a standard, though rather strong axiom of continuity: it would simply say that the graph of the relation $\succsim$ is closed in $X \times X$. This axiom is stronger than the standard continuity axiom of consumer choice, though it is implied by it when the relation $\succsim$ is also known to be a weak order. In our case, however, the consequent of the axiom is only required to hold if the sequences are comparable. As explained above, $x_n \succsim y_n$ for all n may not imply $x \succsim y$ (in the limit) if, for example, y is the only element involved that is in X^0; in this case it can enjoy the extra utility derived from obeying the principle, and thus $y \succ x$ can occur at the limit with no hint of this preference emerging along the sequence.

Clearly, if we restrict attention to one subspace, that is, if all of (x_n), (y_n), x, y are in X^1 or if all of them are in X^0, we obtain a standard continuity condition. Indeed, this would suffice to represent $\succsim$ on X^0 by a continuous

utility function u^0 and to represent it on X^1 by a continuous utility function u^1, where u^0 and u^1 (having disjoint domains) need not have anything in common.

While A2 deals with weak preferences that are carried over to the limit, we will also need a corresponding axiom for strict preferences:

A3. Strict Preference Continuity: For all comparable sequences $x_n \to x$ and $y_n \to y$, and all $z, w \in X$, if $x_n \succsim z \succ w \succsim y_n$ for all n, then $x \succ y$.

To see the meaning of this axiom, assume, again, that comparability were not required. In this case, $x_n \succsim z$ and $w \succsim y_n$ would imply $x \succsim z$ and $w \succsim y$, respectively, and from $z \succ w$ we would easily conclude $x \succ y$. In our case, however, we could have that $(x_n) \subset X^1$ and $x \in X^0$, and thus we cannot conclude that $x \succsim z$ (and, naturally, the same holds for w and y). Yet, comparability of $x_n \to x$ and $y_n \to y$ suffices to conclude that the preference gap between z and w is indeed enough to guarantee a strict preference between x and y.

Next, we introduce an Archimedean axiom stating that the "cost" of the principle in terms of utility is strictly positive, and, moreover, that no utility difference over X^0 exceeds infinitely many such "costs". Specifically, consider a sequence $(z_n) \subset X^1$ that converges to a point $z \in X^0$. In terms of hedonic utility, the bundles z_n become practically indistinguishable from z. However, the fact that z satisfies the principle means that its overall utility is higher than the limit of the corresponding utility values along the sequence. Intuitively, reaching X^0 at the limit provides an extra utility boost, which is not captured by the (continuous) hedonic utility, but should be captured in our overall-utility representation. One way to see this in terms of preferences is the following: if, along the sequence, $z_n \sim y \in X^0$, then we should have strict preference at the limit, $z \succ y$. In this case, the (hedonic) utility gap between z and y is a measure of the contribution of the principle to overall utility. The axiom states that, when aggregated, these measures are large

enough to cover the entire utility range over X^0. Explicitly,

A4 Archimedeanity: Let $\left(x^k, z^k\right) \subset X^0$ and $\left(z_n^k\right)_{n,k\geq 1} \subset X^1$ be such that (i) $z_n^k \to z^k$; (ii) $x^k \succsim z^k$ and (iii) $z_n^k \succsim x^{k+1}$ for all $k \geq 1$ ($z_n^k \succsim x^{k-1}$ for $k \geq 2$). Then there does not exist $\hat{x} \in X$ such that $x^k \succsim \hat{x}$ ($\hat{x} \succsim x^k$) for all $k \geq 1$.

Finally, we find it convenient to rule out the case in which all points in X^0 are equivalent.

A5 Non-Triviality: There are $x, y \in X^0$ such that $x \succ y$.

2.3 Results

We are now ready to state our behavioral characterization of preferences that satisfy the aforementioned axioms.

Theorem 1 *Let there be given $d \in \mathcal{D}$ and $\succsim$. The relation $\succsim$ satisfies A1-A5 if and only if there exist a continuous function $u : X \to \mathbb{R}$, which isn't constant on X^0, and a constant $\gamma > 0$ such that $\succsim$ is represented by*

$$u_d(x) = u(x) - \gamma \mathbf{1}_{\{d \cdot x > 0\}} \tag{7}$$

As discussed in the Introduction, the representation (7) captures an agent whose choices are driven by two factors: on the one hand, the desire to maximize hedonic well-being – measured, as usual, by u – and, on the other hand, the desire to abide by an intrinsic principle – whose violation affects overall well-being by the penalty γ.

The proof appears in Appendix B. To better understand its logic, we first note that axioms A1 and A2 trivially imply that one can find continuous representations of $\succsim$ on X^0 and on X^1, because on each of these A2 implies that standard continuity axiom. This, however, does not mean that there exists a function that is continuous on all of X and that represents $\succsim$ both on X^0 and on X^1 (separately). Appendix A is devoted to an auxiliary result, stating

that A3 is the missing link. The result (formally stated in the appendix) says that any bounded and continuous utility function that represents $\succsim$ on X^1 has a unique continuous extension to X^0, in such a way that the extension represents $\succsim$ also on X^0. Thus, A1-A3 can help us find functions that we can think of as the hedonic utility u above: each is continuous throughout X and correctly represents preferences on each of X^0, X^1. We do not expect any of them to represent preferences across the two spaces, because we know that discontinuities are to be observed between them. More concretely, when a sequence in X^1 converges to a limit point in X^0, we expect the overall utility (u_d) to "jump" in a discontinuous way, where the utility of limit point gets the boost of obeying the principle (while having practically the same hedonic utility u as the tail of the sequence).

The question we turn to is whether, among all the functions u as above, there exist some for which the "boost" in utility when reaching X^0 is a constant $\gamma > 0$. Axiom A4 guarantees that this is possible. First, it guarantees that the boost is strictly positive at any point in X^0. Indeed, axioms A1-A3 do not preclude the possibility that at some points in X^0 the principle is valued (corresponding to $\gamma > 0$) while at others it isn't (as if $\gamma = 0$). A4 does preclude this possibility. More formally, it is easy to see that A4 implies

Discontinuity: Let $x, y, z \in X^0$, and let there be a sequence $z_n \to z$ with $(z_n) \subset X^1$ such that $x \succsim z$ and $z_n \succsim y$. Then $x \succ y$.

As explained in the presentation of A4, we may focus on the case $z_n \sim y$ and $x \sim z$, where standard continuity (over all of X) would imply $x \sim y$, whereas the Discontinuity condition *demands* that a utility gain will be obtained at the limit, to result in strict preference $x \succ y$. Once this is guaranteed, to obtain a representation as in (7) one may exploit the ordinality of standard utility representations and choose to "scale" the utility in such a way that the boost is indeed a constant $\gamma > 0$. This is basically the strategy of the proof: we define "steps" on X^0 that intuitively correspond to "better

than... by exactly the utility of the principle", find a utility function that increases by the same amount for each such step, and extend it to all of X. For this strategy to succeed, we also need to make sure that these steps go far enough. In other words, we wish to make sure that no alternative in X^0 is so much better than any other that the utility difference between them is incommensurable with the steps. And this is the precise meaning of A4.

While A4 is somewhat cumbersome, it is implied by the following axiom:

A6 Lipschitz: There exists $\delta > 0$ such that, for every $x, y, z \in X^0$, and every sequence $z_n \to z$ with $(z_n) \subset X^1$ such that $x \succsim z$ and $z_n \succsim y$, we have $\|x - y\| > \delta$.

Axiom A6 states that, for a bundle $x \in X^0$ to be better than another bundle, $y \in X^0$, by "at least the cost of the principle", x should not be too close to y. We dub it "Lipschitz" as it will be satisfied by any utility function that is Lipschitz continuous on the entire space. Observe, however, that we only require the Lipschitz condition for one specific $\delta > 0$, guaranteeing that two bundles that are δ-close will not have a utility gap that is higher than a certain threshold (the presumed γ). If we restrict attention to compact bundle spaces, we can use A6 in lieu of A4:

Corollary 1 *Let there be given $d \in \mathcal{D}$ and $\succsim$ and assume that X is compact. If the relation $\succsim$ satisfies A1-A3, A5, and A6, there exist a continuous function $u : X \to \mathbb{R}$, which isn't constant on X^0, and a constant $\gamma > 0$ such that $\succsim$ is represented by*

$$u_d(x) = u(x) - \gamma \mathbf{1}_{\{d \cdot x > 0\}}$$

Appendix B also contains the brief proof of this result.

2.4 Observability and Uniqueness

2.4.1 Observability

To what extent are axioms A4 and A6 observable? Evidently, they do not satisfy strict Popperian criteria, because their antecedents involve infinitely many observations (such as, in the case of A4, $x^k \succsim z^k$, $z_n^k \succsim x^{k+1}$ for all n, k). In this sense, A4 and A6 are similar to the standard continuity axiom, as well as to our A2 and A3: requiring infinitely many observations of the type $x_n \succsim y$ in their antecedent, these axioms cannot be refuted by any finite database.

This may suggest that one cannot tell the difference between the axioms discussed here and standard continuity. Mathematically, it is easy to see that our A2 and A3 (even combined) are strictly weaker than full-strength continuity, and, furthermore, that A4 and A6 contradict it (as mentioned above, each implies the *Discontinuity* condition). But since all of these involve infinitely many observations, perhaps it does not really matter which axioms one adopts?

We believe that the answer is negative. There are several scenarios in which one can infer the existence of discontinuities from observed data. Before listing these, observe that in our model the consumer takes into account the product dx. That is, we implicitly assume that the consumer is rational enough to obtain the information about the products, d, and to process it in such a way that only the product dx matters. Under this rationality assumption, discontinuity at $x_i = 0$ can be tested via discontinuity at $d_i = 0$. In other words, instead of observing the consumer's choice between pairs of bundles that vary in minuscule quantities, we can observe her choice given different d vectors. In the following we focus on evidence of discontinuity of demand near a point $d_i x_i = 0$.

First, one may use statistical analysis to extrapolate from finitely many observations to infinitely many. For concreteness, suppose that we can collect data on purchase behavior given different "Nutrition Facts" tables. Assume

that in observation t, the Nutrition Facts declare that the product contains d_{it} grams of meat per serving, and the amount purchased was x_{it}. Assume that all d_{it} values are relatively small, and some are zero. We can run the OLS model $x_{it} = \alpha + bd_{it} + \varepsilon_t$ on the values $d_{it} > 0$ and compare the estimate of the intercept, $\hat{\alpha}$, with the average consumption for $d_{it} = 0$, denoted $\bar{x}_i^0$. If preferences are continuous at zero, one would expect that the linear model would serve as a good approximation and $\hat{\alpha}$ would not differ significantly from $\bar{x}_i^0$. Rejection of the hypothesis that they are equal could be an indication of discontinuity.

Second, one may use imprecise information about the vector d to find indication of discontinuity as in the case of "Nutrition Facts" tables that provide information in terms of ranges. Specifically, it is not uncommon that the information about an ingredient be given as "Less than 1g". The manner in which consumers interpret this piece of information is clearly open to discussion. But it seems reasonable that most consumers who read the label would understand that (i) "Less than 1g" refers to a positive quantity; (ii) this positive quantity is perceived negligible by experts. One can therefore compare consumption given the information "Less than 1g" and given the information "0". Under continuity, one would expect these consumption data to be indistinguishable. Again, a statistically significant difference might be taken as evidence of discontinuity.

Third, one may derive discontinuity analytically given other assumptions. For example, assume that we adopt a condition of monotonicity, stating that standard (weak) monotonicity holds on each of X^0 and X^1 separately. In that case, a violation of monotonicity across the two spaces would have to entail discontinuity. Specifically, if $x \in X^0$ and $y \in X^1$ satisfy $y \geq x$ but $x \succ y$, monotonicity on X^1 implies that $x \succ y \succsim \alpha x + (1-\alpha)\, y$ for all $\alpha \in (0,1)$, in contradiction to continuity.

Lastly, one may have access to preferences beyond the strict interpretation of the revealed preference relation. One class of such preferences include

mind experiments, where a person may use introspection and say "I would not choose this product if it contains any amount of meat". Introspection and stated preferences may not always correspond to actual choices, but they do provide a source of information. Relatedly, preferences might be explicitly stated in order to be communicated. For example, a person might give instructions to a member of the same household to purchase only vegetarian food. When we consider stated preferences, discontinuity is rather natural.[12]

2.4.2 Uniqueness

To what extent is the representation unique? The answer depends on the range of u and on γ. For example, if $\gamma > \sup_{x\in X^1}(u(x)) - \inf_{x\in X^0}(u(x))$, we have $u_d(x) > u_d(y)$ for all $x \in X^0, y \in X^1$ and the consumer would never give up the principle. In this case the utility function is only ordinal: any monotone transformation of u and γ that satisfies the above inequality represents preferences, and the utility function is far from unique. If, by contrast, γ is very small relative to $\sup_{x\in X^1}(u(x)) - \inf_{x\in X^0}(u(x)) > 0$, the monotone transformations that respect the representation (7) are much more limited. As will be clear from the proof, one can choose u more or less freely until a point of equivalence between two bundles $x \in X^0, y \in X^1$, and then the utility is uniquely determined throughout the preference-overlap between X^0 and X^1. Clearly, shifting u by a constant and multiplying both u and γ by a positive constant is always possible. Thus, on the preference-overlap between X^0 and X^1 we have a cardinal representation, and outside this preference interval – only an ordinal one.[13]

[12]Rubinstein (1988) introduced the notion of *definable preferences*, and called for modeling preferences that can be described within a formal language. For example, the lexicographic order, which might appear as a mathematical anomaly when using calculus, is a rather natural example when preferences are stated in natural language.

[13]This is reminiscent of the degree of uniqueness of representations of a semi-order by a function u and a just-noticeable-difference $\delta > 0$. See, for instance, Beja and Gilboa (1992).

3 Extensions

3.1 Other Subspaces

There are situations in which a principle is satisfied on a subspace of alternatives which is not necessarily on the boundary of the entire space. Consider the following example. A parent writes a will. She has two children who are twins, and she cherishes equality. It might be important to her to know that she has behaved fairly in her bequest, and she might also think about the emotional effect that unfair division might have on her children. Denoting by x_i the proportion of the estate bequeathed to child i, the parent can choose any point in $[0,1]^2$ subject to the budget constraint $x_1 + x_2 \leq 1$. Given her preference for equality, she might prefer the point $(0.48, 0.48)$ to $(0.49, 0.51)$, violating monotonicity. Indeed, she is expected to violate continuity near the diagonal $x_1 = x_2$.

This set up isn't a special case of our theorem, but the theorem can easily be adapted to include it. First, we need to allow the vector d to assume values beyond $\{0,1\}$. Then we can describe the space in which the principle holds, X^0, by $d \cdot x = 0$ where $d = (-1, 1)$. Second, in this example the complement of X^0, X^1, is not convex and not even connected. However, it is the union of two convex sets. This means that we can define the utility on each subspace of X^1 separately, and a similar result would hold. In particular, with the necessary adjustments and an additional symmetry axiom, one can obtain a representation of the parent's preferences by a utility function

$$u_d(x_1, x_2) = u(x_1, x_2) - \gamma \mathbf{1}_{\{x_1 \neq x_2\}}$$

where u is continuous. While Ben-Porath and Gilboa (1994) and Fehr and Schmidt (1999) treat inequality continuously, the present formulation allows for discontinuity, conceptualizing equality as an intrinsic value.

3.2 Variable Information

In our analysis above the consumer is assumed to have information about which goods satisfy the principle, embodied in the vector d. In fact, this information could also be revealed from preferences: the space X is divided into X^0 and X^1, with continuity holding on each of these but failing to hold at each point of X^0 when approached by points in X^1. Thus, preferences contain sufficient information to identify X^0, and it can be easily checked whether X^0 is defined by $d \cdot x = 0$ for some indicator vector d.

However, more generally, we may be interested in preferences given different vectors d. Thus, a natural extension of the analysis will be to consider a set of relation $\{\succsim_d\}_d$, one for each possible indicator vector d, and seek a joint representation

$$u_d(x) = u(x) - \gamma \mathbf{1}_{\{d \cdot x > 0\}}$$

with the same (u, γ) that apply to all d.

3.3 Multiple Principles

It is not uncommon for economic agents to have more than one principle. In verbal discussions people tend to espouse many principles, each of which sounds convincing on its own. The question then arises, what will they do when these principles are in conflict with each other and/or with hedonic well-being? Suppose that a vegan consumer also cares about fair trade practices. What would be her choices if, on a supermarket shelf, there are no products that satisfy both principles? Will she choose to eat non-vegan products, vegan that failed to respect fair trade practices, or to skip a meal?

We suggest to model these choices along similar lines, using a utility function that takes into account all principles involved, as well as hedonic well-being. Specifically, assume that there are m principles, denoted by $M = \{1, ..., m\}$, and that preferences are parametrized by a matrix $D = (d_{ij})_{i \leq n, j \leq m}$ such that $d_{ij} \in \{0, 1\}$ denotes whether product i violates

principle j. That is, the consumer is assumed to know which product satisfies which principles. Again, it is assumed that producers are required to mark their products truthfully. We postulate an additive form that generalizes (6). First, given a matrix D, let D^j be its j-th column, so that $(D^j)_i = D_{ij}$. Next, assume that for each principle j there exists $\gamma_j > 0$ such that, given the matrix D, the consumer maximizes

$$u_D(x) = u(x) - \sum_{j=1}^{m} \gamma_j \mathbf{1}_{\{x \cdot D^j > 0\}}$$

where u is continuous.

3.4 Instrumental Values

Instrumental values are means rather than ends. As in the example of CO_2 emission, agents care about them because they are understood to affect the values one inherently cherishes. Typically there exists some mechanism that underlies the relationship between the instrumental and the intrinsic value that is ultimately behind it. The mechanism can be physical, chemical, or biological, as in the example of the effect of CO_2 emission on global warming, and on wildlife preservation. Sometimes an economic or social mechanism is involved. For example, affirmative action is often justified based on its long term effects through role models. Be that as it may, mechanisms tend to be continuous. An agent who wishes to minimize global warming will not care about a few grams of CO_2 emitted by a flight in the same way that a vegetarian would care about a few grams of meat in her plate; similarly, an agent who wishes to support minority groups role models because of their long run effects on equality would tend to think of the value in a more continuous way than one thinks about a just bequest (as in Subsection 3.1).

We are therefore led to model instrumental values by

$$u_d(x) = u(x) + v(d, x)$$

where $v(d, x)$ is a continuous function. It is also natural to allow $d \in \mathbb{R}^n$ to assume values beyond $\{0, 1\}$, and to represent the *degree* to which products hurt the value in question. It stands to reason that $v(d, x)$ will only depend on $d \cdot x$. For example, if the production and consumption of a unit of good i cause the emission of d_i grams of CO_2 into the atmosphere, the total emission of a bundle x is $d \cdot x$ and its effect on the agent's utility is $v(d, x) = \hat{v}(d \cdot x)$.

Observe that a typical utility function u_d would now be continuous (for each d) but not monotone. We do not axiomatize such functions here. However, we illustrate the model by a simple example, paralleling the example in Section 1.3.

3.4.1 An Example

An agent has to decide how much to travel by air. We can think of a simplified model with two consumption goods: let x_1 denote the quantity of flights consumed, and x_2 – the quantity of an aggregate good. This aggregate good contains complementary goods, such as rail travel, as well as other, unrelated goods. Let the prices per unit be p_1 and p_2, respectively, and let I denote the consumer's income. Let us first assume that the agent's utility function is a standard Cobb-Douglas utility

$$u(x_1, x_2) = \alpha \log(x_1) + (1 - \alpha) \log(x_2)$$

(for $\alpha \in (0, 1)$) so that the consumer's expenditure on flights will be αI.

Next assume that the consumer cares not only about her hedonic well-being, but also about the emission of CO_2: she suffers disutility from the knowledge that her consumption causes damage to the environment. Assume that each unit of air travel, x_1, hurts the environment to degree $d > 0$, and that the consumer cares about this damage to degree $\gamma \geq 0$. We will now assume that, *given* the value of d, the consumer maximizes

$$u_d(x_1, x_2) = \alpha \log(x_1) + (1 - \alpha) \log(x_2) - \gamma d x_1$$

subject to the same (standard) budget constraint. Observe that the last term of u_d – the disutility caused by the knowledge of environmental damage – is linear in x_1. In this formulation we do not assume a decreasing marginal disutility pattern, because this component of the utility isn't perceived physiologically, nor does it follow from the degree to which various needs are satisfied. Rather, it is the negative impact of a purely cognitive phenomenon, namely, awareness of the impact one's consumption has.

Note that the function u_d is increasing in x_2 throughout the range, but it is increasing in x_1 only in the region $x_1 \leq \frac{\alpha}{\gamma d}$. However, these preferences are convex. For any value of $\gamma > 0$ and all positive (p_1, p_2, I), the optimal solution will be obtained in the range $x_1 < \frac{\alpha}{\gamma d}$, and it will be an interior solution satisfying

$$x_1 = \frac{1}{2}\left[\frac{I}{p_1} + \frac{1}{\gamma d} - \sqrt{\left(\frac{I}{p_1} + \frac{1}{\gamma d}\right)^2 - \frac{4\alpha I}{\gamma d p_1}}\right]$$

Thus, if we can observe the consumer's choice given the standard parameters (p_1, p_2, I) as well as the new parameter d, we can solve for both α and γ. Indeed, in this simple example, the consumer's choice for $d = 0$ is sufficient to derive α, and one more observation of the optimal choice for some $d > 0$ is sufficient, in principle, to factor out γ.

3.5 Combined Models

Intrinsic values that exhibit discontinuity at zero might also be strictly increasing beyond zero. For example, a vegetarian consumer may not only categorize foods as "vegetarian" or "non-vegetarian", and might care about the number of animals and/or the species of animals that had to be sacrificed for the meal. It is therefore natural to think of a function $u_d(x) = u(x) + v(d, x)$ where v is not dichotomous, yet discontinuous at zero. Given such a general framework, one can pose the question: is a given value intrinsic or instrumental? We hold that question might be relevant to public policy. Policy

should take values into account, but it should do so differently depending on whether the values in question are intrinsic or not: instrumental values can be replaced by other means to achieve the goals that truly matter, while intrinsic values cannot be negotiated. Consider the following example. Suppose that people express a preference for the preservation of lizards. Some might think that lizards are as cute as kittens, and have a right to live peacefully as do chimpanzees or dolphins. That is, for some people the preservation of lizards is an intrinsic value. Others might think that lizards are very useful as they keep spiders away. This is evidently a more instrumental approach to lizards. If the majority of society is of the latter type, we may not insist on preserving lizards in case there are other solutions to the spider problem. But if most people do feel that lizards are in the same category as are kittens, it seems pointless to suggest to them alternative solutions to the benefits derived from lizards. Testing whether preferences for preservation of lizards are continuous near zero might help us in determining which is the case.

4 Related Literature

It has long been observed that consumers care about ethical values. Auger, Burke, Devinney, and Louviere (2003) and Prasad, Kimeldorf, Meyer, and Robinson (2004) found that consumers were conscientious and expressed willingness to pay more for products that had desirable social features, such as environmental protectionism, avoiding child labor, as well as sweatshops. Barnett, Cloke, Clarke, and Malpass (2005) discussed the notion of "consuming ethics". As mentioned above, De Pelsmacker, Driesen, and Rayp (2005) estimated the willingness to pay for coffee that was and wasn't labeled as "Fair Trade" and found significant differences, with some (about 10%) of the sampled consumer willing to pay a premium that was 27% for the label. Loureiro and Lotade (2005) found similar results for Fair Trade and Eco labels, and Basu and Hicks (2008) – for Fair Trade coffee in a cross-national

study. Enax, Krapp, Piehl, and Weber (2015) found neurological evidence for the positive effects of social sustainability. Arnot, Boxall, and Cash (2006) used revealed preference data and found a significant effect, with lower price sensitivity in the labeled product as compared to the unlabeled one. More recently, Hainmueller, Hiscox, and Sequeira (2015) conducted a study in which they collected actual purchase data and showed that the "Fair Trade" label increased sales by 10%. The standard methodology in these studies is discrete choice modeling, where a random utility model is estimated, and the effect of a label can be tested. These estimations can present the same product with different labels (in our language, compare (d, x) with (d', x)). Our approach can be viewed as seeking to provide axiomatic foundations for these works, with a focus on cases in which one cannot credibly attach different labels (such as "vegetarian"/"non-vegetarian") to the same good.

While the above works focus only on the consumer side, Bartling, Weber, and Yao (2015) and Pigors and Rockenbach (2016) study market behavior in laboratory experiments and, in particular, the role of consumers in inducing socially responsible production. In Bartling, Weber, and Yao (2015), consumers can choose between two types of products that are equivalent in terms of hedonic well-being, and that differ only in the dimension of social responsibility—the "unfair"product costs less to produce but generates a negative externality to a third party. Contrary to the standard equilibrium prediction, the authors show that a significant fraction of consumers are willing to pay a price premium for the 'fair' product, producers are willing to share the burden by incurring higher production costs, and such willingness persists under repeated market interactions. Pigors and Rockenbach (2016) argue that firm competition is beneficial for the emergence of socially responsible production. Their experiment shows that in monopolistic conditions, consumers are more price-sensitive and socially responsible monopolists suffer a decrease in profits. In the presence of competition, however, consumers are willing to pay higher prices for socially responsible products, leading to

significantly higher profits for producers.

Taking a broader perspective, the notion that consumption has socio-psychological effects has long been recognized. Veblen (1899) suggested the notion of conspicuous consumption, and Duesenberry (1949) formulated the relative income hypothesis, both having to do with determinant of well-being that go beyond the physical. Frank (1985a, 1985b) highlighted the role of social status, and, more recently, Heffetz (2011) studied the effects of conspicuous consumption empirically. Interdependent preferences are also at the core of Fehr and Schmidt's (1999) inequity aversion, Karni and Safra's (2002) sense of justice, as well as Ben-Porath and Gilboa's (1994) axiomatization of the Gini Index, and Maccheroni, Marinacci, and Rustichini's (2012) model, which is applied in Maccheroni, Marinacci, and Rustichini's (2014) to show the economic effects of envy and pride. However, values, and more generally the *meaning* of consumption, seem to be understudied in economics. Conspicuous consumption can be viewed as dealing with meaning, reflecting on one's social standing, and thereby on one's identity. Inequity aversion can similarly be conceived of as an attitude towards the value of equality. But meaning and values that are not related to social ranking are typically neglected in formal, general-purpose models of utility.

The more applied economic literature has addressed specific values more directly and explicitly. For example, Barbier (1993) and Morrison (2002) study use and non-use values of wetlands. Barnes, Schier, and van Rooy (1997) examine the value of wildlife preservation, while Bedate, Herrero, and Sanz (2004) – of cultural heritage. Hornsten and Fredman (2000) and Chen and Qi (2018) deal with the value attached to forests in or near urban areas. Most of this literature relies on the Contingent Valuation Method (CVM), which is based on self-reported willingness to pay. Throsby (2003) discusses this measure and criticizes it. Given this criticism, and psychological findings such as Kahneman and Knetch (1992), one may be wary of CVM findings. Indeed, Bedate, Herrero, and Sanz (2004) adopt an idea suggested by Hotelling

(1947), to use travel time as a way to measure the value of cultural heritage. This is indeed a measure that relies on economic choices rather than on (often hypothetical) self-report, but it cannot apply to many values in question. Even a related example such as the preservation of species in the depth of the oceans cannot be measured by travel time decisions. Importantly, given that economic theory cherishes revealed preferences and tends to dismiss verbal self-reports as an unreliable source of data, most of the values discussed in this literature hardly play any explicit role in microeconomic theory models. The rational consumer whose preferences are described by a preference relation $\succsim$ seems to care about quantities of goods, and not about what they signify.

This view of economic agents has been criticized as yet another feature of *homo economicus*, the much-ridiculed fictional character whose sole habitat, allegedly, is economic models. Medin, Schwartz, Blok, and Birnbaum (1999) argued against formal models in economics and decision theory precisely on these grounds, namely, that these models do not pay attention to meaning and signification. According to their approach, decision theory lacks the *semantics* of decisions. In various questionnaires they showed rather intuitive results about meaning of actions. For example, many participants in their experiments reported that they would not sell their wedding ring for any material payoff, but they would do so to save their child. Similarly, the amount of money they would demand for a real estate property would not depend only on its economic worth, but also on how long it has been in family possession. In both examples, meaning is key. A wedding ring isn't just a piece of gold; it signifies love and devotion. It should be priceless when "price" is measured in money, but it can be sacrificed to save the life of a beloved joint child. Saving the life of a child would endow the sale with meaning that no material consumption can generate. Along similar lines, a family property can mean a lot to the family members, in ways that the market value would not reflect.

The literature in marketing deals with meaning and signification of goods. A large and vibrant field of research asks what goods mean to consumers and what values they signify (see Sheth, Newman, and Gross, 1991). Moreover, goods are sometimes perceived as determinants of consumer's identity. In particular, some of the explanations of brand loyalty, especially in the context of upscale brand names, involve identity. A consumer might think of himself as the "kind of person who wears...", where the good clearly becomes more than a physical product that satisfies some needs (see He, Li, and Harris, 2012). Moreover, Consumer Culture Theory is, to a large extent, about what consumption means, and not about what it is as a mere economic activity. (See Arnould and Thompson, 2005, and, more recently, Bajde, 2014.) However, the analysis in these strands of the literature usually does not involve formal modeling in a way that can be incorporated into microeconomic theory. Whether the analysis is qualitative in nature, or focuses on experimental and empirical data, it does not suggest to an economist a model that can replace the standard model of neoclassical utility maximization. Indeed, Calabresi (1985, 2014) discusses this point in the context of law and economics, and the degree to which economic models can capture the values society cares about.

Recent developments in behavioral economics suggested formal modeling of some related phenomena. Dillenberger and Sadowski (2012) and Evren and Minardi (2015) model and axiomatically derive affective responses to the ethical judgment of one's choices. The former deal with shame over selfish behavior, and the latter – with the "warm glow" effect, namely, the positive affective response to having made an ethical choice. These works are similar to ours in introducing ethical considerations into the utility function. They differ in terms of the set-up and assumptions (using menu choices and continuous preferences). We return to discuss warm glow effects in subsection 5.5 below.

Meaning is also related to narratives, to stories one can construct. In-

deed, Eliaz and Spiegler (2018) deal with narratives of causality, and Glazer and Rubinstein (2020) – with stories that are sequences of events. However, both deal with narratives as constructions of beliefs, whereas our focus here is on their role as determinants of utility. There have been studies that challenge this dichotomy: Brunnermeier and Parker (2005) and Bracha and Brown (2012) model agents who choose not only what to do, but also what to believe (under certain constraints). The agents we aim to model, by contrast, accept information as given. It is implicitly assumed to be truthful, and, while it may factor into the agents' sense of identity and well-being, we do not assume that they choose what to believe or even how to interpret that information. For example, a vegetarian accepts information about the ingredients of food products, and we wish to study how this information changes her consumption behavior via the value of vegetarianism, but without more involved processes such as constructing narratives or choosing what to believe.

5 Discussion

5.1 Incomplete Information

As briefly discussed in the Introduction, one may wonder whether we need a model in which values feature explicitly, given that the neoclassical utility function is derived from observed choices. An alternative approach would be to use the standard model, and in case there is some information that is relevant to consumption – that is, the vector d – to view it as an incomplete information model. For example, in our model of Section 2 we could think of a food product as being known to be vegetarian, known to be non-vegetarian, or not known to be either. The latter could be viewed as a state of uncertainty, where the consumer has two states of the world in mind, and, as long as the good's classification in unknown, considers its expected utility.

This approach is certainly possible, and under certain conditions one

could attempt to elicit (i) the consumer's utility for the good in case it is known to be vegetarian or not, and (ii) the consumer's subjective probability of each state. Indeed, one can view our approach as eliciting (i) without (ii). However, the standard assumptions of expected utility theory – and of many variants and generalizations thereof – may not hold in this case. A consumer who has a preference for vegetarianism will typically violate consequentialism:[14] her utility is partly determined by the knowledge that she has or has not respected the value, in a way that isn't captured by the observable properties of her bundle. Such a consumer may devote resources to find out whether the food she has consumed in the past was vegetarian; and she may care about vegetarianism more or less in the future if she finds out that she has betrayed this principle in the past. Finally, because the knowledge that she has – or has not – respected a principle in the past factors into her utility, such a consumer will also not be indifferent to the timing of resolution of uncertainty. To sum, a formal model of utility given different information states about the products cannot be simply derived from a Bayesian model of consumption under uncertainty.

5.2 Other Distinctions

Our discussion only distinguishes between intrinsic and instrumental values, and attempts to map this conceptual distinction to the question of (dis)continuity at zero. There are, however, several other distinctions that may be conceptually insightful. First, let us revisit the production/consumption distinction. If a consumer happens to obtain a good as a gift, she might have lesser guilt feelings having to do with its production, as she has not chosen to buy it, and has not spent money to support the industry. By contrast, guilt feelings that have to do with its consumption are unaffected by the origin of

[14] As is usually the case, one can salvage consequentialism by introducing the knowledge of one's past consumption, and the matrix D, into the notion of "a consequence". This exercise is always possible, and precisely for that reason, it renders consequentialism vacuous.

the good.

Another distinction has to do with the degree to which values are translated to utilitarian calculations, and the degree that generalizations are required for the exercise. For example, a consumer might feel that she is too small to affect the number of flights, and thus, with a negligible marginal contribution, she need not feel bad about boarding a plane that is "anyway" taking off. To explain the sense of moral responsibility of such a consumer we might need to resort to a Categorical Imperative type of argument, as in the case of voting. By contrast, when a person fishes a fish for lunch, she can more directly see the consequence of her actions.

While such distinctions might, at least under some circumstances, be observable, we do not study them in this paper. Thus, we equate purchase with consumption, and do not delve into the psychological origins of the negative feelings caused by compromising values.

5.3 Unawareness

Our model assumes that the vector d is known, and, in particular, that the agent is fully aware of it. Our agents can therefore be fully rational, (provided that we do not rule out morality and values as "irrational"). We therefore assume that the values of d_i are reported whether they are positive or zero, so that there is no question of awareness of the principle, nor of uncertainty about d_i. One may extend the model to allow for the possibility that d_i isn't reported at all. This can capture a wider range of phenomena. For example, an agent who is about to take a flight might not be thinking about its environmental effects. Once airlines start reporting the environmental damage per flight (d_i) – the agent may suddenly be aware of the value-effect of her consumption decisions, and perhaps change them.

5.4 Meaning and Well-Being

The literature on well-being recognizes that it has both hedonic and eudaimonic determinants. The former refers to the instantaneous positive and negative sensations, whereas the latter – to a sense of meaning, self-fulfilment, and so forth. (For a review see Ryan, 2001.) Our model can be viewed as dealing with these factors as well. For example, consider a person who wishes to give his children broad cultural education, and, to reach this goal, is willing to give up hedonic well-being, commute longer time to work etc. We could view this person as deriving well-being from the meaning that his material sacrifice has. We could also think of him as having a *value* of enriching his children's education. Indeed, while "values" have moral connotations, in some cases it may be hard to judge whether certain cognitions are values or otherwise imbue life with meaning.

5.5 Donations

Extended versions of our model can also be used to describe the choice of donations. A donation could be thought of as a good x_i that does not affect the function u but that enters the function v in a way that increases well-being; that is, $u(x)$ is independent of x_i but $v(d, x)$ is increasing in it. The price for monetary donations would naturally be $p_i = 1$, and the information state d should describe what causes are served by the donated amount. In this way, the "warm glow" of donations is introduced into the utility function, but, as opposed to Evren and Minardi (2015), in this model the extent of its effect on well-being is not determined by the available menu of choices.

What form would the function $v(d, x)$ take in the case of donations? We would surely expect it to be strictly increasing in the donated amount x_i. It is less obvious whether it should be continuous at zero. On the one hand, a rational consumer should realize that donating for a cause is basically supporting an instrumental value. The act of donation itself is only a transfer of a sum of money between bank accounts, and it is hard to ascribe profound

meaning to this act per se. Rather, it is the ultimate goal that this money will help support that is a carrier of meaning. The mechanism by which one's money is translated to, say, feeding hungry children, introduces continuity. On the other hand, some feeling of warm glow might result from very small amounts as well. Indeed, fund raisers might ask for a contribution, "no matter how small". And any positive donation allows one to truthfully say – to others as well as to oneself – that one has donated money. Finally, faith and religious sentiments might endow a donation with positive meaning in a way that is, to a large extent, detached from the amount donated. We thus see room both for continuous and discontinuous models of donations.

6 Appendix A: An Auxiliary Result

In this appendix, we present and prove the following result.

Theorem 2 *Let $\succsim$ on X satisfy A1-A3. Then, a bounded and continuous function $u : X^1 \to \mathbb{R}$ that represents $\succsim$ on X^1 has a unique continuous extension to (all of) X. This extension represents $\succsim$ also on X^0.*

Note that the theorem does not state that the extended u represents $\succsim$ on X in its entirety. Indeed, the continuity axioms do not state that preferences change continuously along a sequence that crosses from X^1 to X^0, and thus a utility function that is continuous on the entire space cannot be expected to represent preferences across the two subspaces.

6.1 Proof of Theorem 2

Without loss of generality we assume that d isn't identically 0 not identically 1, so that $X^0, X^1 \neq \varnothing$. Note that, due to convexity of X^1, X^0 is included in the closure of X^1.

We start with a few lemmas. Throughout we assume that $\succsim$ on X satisfies A1-A3. (Note, however, that the first three lemmas do not make use of A3).

Lemma 1 Let there be a sequence $x_n \to x$. Assume that $[(x_n) \subset X^0$ and $x \in X^0]$ or $[(x_n) \subset X^1$ and $x \in X^1]$. Then, for all $y \in X$, if $x_n \succsim y$, then $x \succsim y$ and if $y \succsim x_n$, then $y \succsim x$.

Proof: Define $y_n = y$ for all $n \geq 1$. Note that the sequences $x_n \to x$ and $y_n \to y$ are comparable (satisfying Condition A), and apply A2. □

Lemma 2 Let there be $x, y, z \in X$ with $x \succ y \succ z$. Assume that $x, z \in X^0$ or that $x, z \in X^1$. Then there exists $\alpha \in [0, 1]$ such that $y \sim \alpha x + (1 - \alpha) z$.

Proof: The argument is familiar, and we mention it explicitly to point out that it does not depend on monotonicity or openness conditions. Let there be $x, y, z \in X$ with $x \succ y \succ z$ and assume without loss of generality that $x, z \in X^0$ (the argument is identical for X^1). Define

$$\begin{aligned} A^- &= \{\, \alpha \in [0,1] \mid y \succ \alpha x + (1-\alpha)\, z \,\} \\ A^+ &= \{\, \alpha \in [0,1] \mid \alpha x + (1-\alpha)\, z \succ y \,\} \end{aligned}$$

and we have $A^- \cap A^+ = \varnothing$, with $1 \in A^+$ and $0 \in A^-$. Consider $\alpha^* = \inf A^+$ and define $x^* = \alpha^* x + (1 - \alpha^*) z$. We wish to show that it is the desired α, so that $\alpha^* \notin A^- \cup A^+$ and $y \sim x^*$ holds. Suppose that this is not the case. If $\alpha^* \in A^-$ (and $y \succ x^*$), we can choose a sequence $\alpha_n^+ \in A^+$ with $\alpha_n^+ \searrow \alpha^*$. Then $x_n = \alpha_n^+ x + (1 - \alpha_n^+) z \in A^+ \to x^*$. Importantly, X^0 is convex. Hence $x_n \in X^0$ for all n and $x^* \in X^0$ as well. Lemma 1 implies that $x^* \succsim y$, a contradiction. Similarly, if $\alpha^* \in A^+$ (and $x^* \succ y$), then $\alpha^* = \min A^+$ and we must have $\alpha^* > 0$ as $0 \in A^-$, in which case we can choose a sequence $\alpha_n^- \in A^-$ with $\alpha_n^- \nearrow \alpha^*$. Then, Lemma 1 implies that $y \succsim x^*$, again a contradiction. Hence $y \sim x^*$.

Note that the argument holds also for X^1 because it is a convex set as well. □

We also note the following.

Lemma 3 For all comparable sequences $\xi_n \to \xi$ and $\eta_n \to \eta$, if $\xi \succ \eta$, then there exists an $N > 0$ such that

$$\xi_n \succ \eta_m \quad \forall n, m > N.$$

Proof: If the conclusion does not hold, for $N_1 = 1$ we have n_1, m_1 such that $\eta_{m_1} \succcurlyeq \xi_{n_1}$. Set $N_2 = \max(n_1, m_1)$ and find $n_2, m_2 > N_2$ such that $\eta_{m_2} \succcurlyeq \xi_{n_2}$. Continuing this way, we generate two subsequences $(n_k, m_k)_k$ such that $\eta_{m_k} \succcurlyeq \xi_{n_k}$ for all k, with $\xi_{n_k} \to \xi$ and $\eta_{m_k} \to \eta$ being comparable (as subsequences of comparable sequences with these limits). A2 would then imply $\eta \succcurlyeq \xi$, a contradiction. □

Two implications of the A3 (in the presence of A1, A2) will be useful to state explicitly.

Lemma 4 For all comparable sequences $x_n \to x$ and $y_n \to x$, and all $z, w \in X$, if ($x_n \succsim z$ and $w \succsim y_n$) then $w \succsim z$.

Proof: Let there be given comparable sequences $x_n \to x$ and $y_n \to x$ as well as $z, w \in X$ such that $x_n \succsim z$ and $w \succsim y_n$. We need to show that $w \succsim z$. Assume, to the contrary, that $z \succ w$. Define $y = x$. With $x_n \succsim z \succ w \succsim y_n$ we can apply A3 and conclude that $x \succ y$ which is impossible as $y = x$. Thus we rule out the possibility $z \succ w$ and conclude that $w \succsim z$ as required. □

The following lemma is not needed for Theorem 2 but will be used later on in the proof of Theorem 1. It is similar to A3 and can easily be shown to imply it. Thus the lemma shows that, in the presence of A1 and A2, the two conditions are equivalent.

Lemma 5 For all pairs of comparable sequences, ($x_n \to x$ and $y_n \to y$) and ($z_n \to z$ and $w_n \to w$), if (i) $z \succ w$; and (ii) $x_n \succsim z_n$; $w_n \succsim y_n$ for all n, then $x \succ y$.

Proof: Assume, then, that ($x_n \to x$ and $y_n \to y$) and ($z_n \to z$ and $w_n \to w$), are given, such that (i) $z \succ w$; and (ii) $x_n \succsim z_n$; $w_n \succsim y_n$ for all

n. We split the argument depending on the reason that $z_n \to z$ and $w_n \to w$ are comparable. Assume, first, that they satisfy Condition A, that is, that $(z_n) \subset X^i, z \in X^i$ and $(w_n) \subset X^j, w \in X^j$ for $i, j \in \{0, 1\}$. In this case, because the limit of each sequence $(z_n), (w_n)$ belongs to the same space X^i as the sequence itself, we also have, w.l.o.g., $z_n \succ w$ and $z \succ w_n$ for all n. (Otherwise, we can apply A2 to the relevant sequence and to a constant sequence and derive $w \succsim z$ from A2.) Next, consider a specific $n > N$. If there are infinitely many indices $n_k > n$ such that $z_{n_k} \succsim z_n$, let n be the minimal index with this property, and, for that n, set $z^* = z_n$ and restrict attention to the subsequence $(n_k)_k$. Clearly, $x_{n_k} \succsim z_{n_k} \succsim z_n = z^*$. If not, then for every $n > N$ there is $l_n > 0$ such that, for all $m > n + l_n$, we have $z_n \succ z_m$. In that case we can select a subsequence (z_{n_k}) such that $z_{n_k} \succ z_{n_{k+1}}$. As $(z_{n_k}) \to z$ and belongs to the same space (as z), we can compare it to the sequence that equals z throughout and conclude that $z_{n_k} \succsim z$ for all k. We can then set $z^* = z$ and we have $x_{n_k} \succsim z_{n_k} \succsim z = z^*$. Thus we found an element z^* and a subsequence (n_k) such that $x_{n_k} \succsim z^*$ with z^* being either z or one of z_n.

We now limit attention to the subsequence (n_k) and repeat the argument for (w_n). In a symmetric fashion, we now have a sub-subsequence (n_{k_l}) and w^* which is either w or one of w_{n_k} such that $w^* \succsim w_{n_{k_l}} \succsim y_{n_{k_l}}$. Importantly, whether $z^* = z_n$ or $z^* = z$, whether $w^* = w_{n_k}$ or $w^* = w$, we have $z^* \succ w^*$ (where this follows either from $z \succ w$, which was given, or from the claims proven above for the other three possibilities). Thus A3 can be used to derive the conclusion $x \succ y$.

Next assume that $z_n \to z$ and $w_n \to w$ are comparable but that they do not satisfy Condition A. This means that the satisfy Condition B, that is, that $(z_n) \subset X^i, z \in X^j$ and $(w_n) \subset X^i, w \in X^j$ for $i, j \in \{0, 1\}$. But this also means that $i \neq j$ (or else Condition A would also hold). Further, because X^0 is closed, we have to have $(z_n), (w_n) \subset X^1$ while $z, w \in X^0$. As X^0 is convex, hence connected, we have $z' \in X^0$ such that $z \succ z' \succ w$ (otherwise, we could

use Lemma 1, applied to $z_n \to z$ and $w_n = w$ to get $w \succsim z$). Repeating the argument for the pair $z' \succ w$, we conclude that there is also $w' \in X^0$ such that

$$z \succ z' \succ w' \succ w.$$

Next we select elements $(z'_n), (w'_n) \subset X^1$ such that $z'_n \to z'$ and $w'_n \to w'$. Notice that this is possible as X^0 is a non-trivial subspace of X. Thus we have four sequences, $z_n \to z$, $w_n = w$, $z'_n \to z'$, $w'_n \to w'$ and two of which are comparable. Applying Lemma 3 consecutively, we conclude that there exists an $N > 1$ such that, for all $n, k, l, m > N$ we have

$$z_n \succ z'_k \succ w'_l \succ w_m.$$

Fix $k, l > N$ and set $z^* = z'_k$, $w^* = w'_l$. Thus, $z_n \succ z^* \succ w^* \succ w_n$ for all $n > N$. As we also have $x_n \succcurlyeq z_n$ and $w_n \succcurlyeq y_n$ for all n, we conclude that $x_n \succ z^* \succ w^* \succ y_n$ and apply A3 to conclude that $x \succ y$. □

We now turn to define the extension. Let there be given a bounded and continuous function $u : X^1 \to \mathbb{R}$ that represents $\succsim$ on X^1. We first note that

Lemma 6 Assume that $(x_n) \subset X^1$ is such that $x_n \to y \in X^0$. Then $\exists \lim_{n\to\infty} u(x_n)$.

Proof: Assume that $x_n \to y \in X^0$. We claim that there exists $a \in \mathbb{R}$ such that $u(x_n) \to a$. If $u(x_n) \to \sup_{x\in X^1} u(x)$ or $u(x_n) \to \inf_{x\in X^1} u(x)$ then $u(x_n)$ is convergent and we are done. Assume, then, that this is not the case. As u is bounded, we can find a number $a \in (\inf_{x\in X^1} u(x), \sup_{x\in X^1} u(x))$ and a subsequence $(x_{n_k})_k$ such that $u(x_{n_k}) \to_{k\to\infty} a$. If we also have $u(x_n) \to_{n\to\infty} a$, we are done. Otherwise, there exists $\varepsilon > 0$ such that, for infinitely many n's, $u(x_n) > a + \varepsilon$, or that, for infinitely many n's, $u(x_n) < a - \varepsilon$ (or both). This means that there is another subsequence $(x_{n_l})_l$ such that $u(x_{n_l}) \to_{l\to\infty} b$ with $|a - b| \geq \varepsilon$. Assume w.l.o.g. that $b \geq a + \varepsilon$. As u is continuous on X^1, and the latter is convex (and connected), we have points $z, w \in X^1$ such that $b - \frac{\varepsilon}{3} > u(z) > u(w) > a + \frac{\varepsilon}{3}$. But this means that, for large enough k, l,

we have $x_{n_l} \succ z \succ w \succ x_{n_k}$ with $x_{n_k} \to_{k\to\infty} y$ and $x_{n_l} \to_{l\to\infty} y$. By A3 we should get $y \succ y$, a contradiction. Thus $u(x_n)$ is convergent. $\square$

Lemma 7 For every $y \in X^0$ there exists $a \in \mathbb{R}$ such that, for every $(x_n) \subset X^1$ with $x_n \to y$, we have $\exists \lim_{n\to\infty} u(x_n) = a$.

Proof: Lemma 6 already established that every convergent sequence $x_n \to y \in X^0$ generates a convergent sequence of utilities. Clearly, this means that the limit is independent of the sequence. Explicitly, if $(x_n), (x'_n) \subset X^1$ are such that $x_n \to y \in X^0$ and $x'_n \to y$, we know that for some $a, a' \in \mathbb{R}$ we have $u(x_n) \to a$ and $u(x'_n) \to a'$. But if $a \neq a'$, we can generate a combined sequence whose utility has no limit. (Say, for $z_{2n} = x_n, z_{2n+1} = x'_n$, we get $z_n \to y$ but $u(z_n)$ is not convergent.) $\square$

We can finally define the extension of u. For every $y \in X^0$ there exist sequences $(x_n) \subset X^1$ with $x_n \to y$. By Lemma 6 we have $\exists \lim_{n\to\infty} u(x_n)$ and by Lemma 7 its value is independent of the choice of the convergent sequence. Thus, setting

$$u(y) = \lim_{n\to\infty} u(x_n)$$

is well-defined. Observe that this is the unique extension of u to X^0 that holds a promise of continuity.

Lemma 8 u is continuous (also) on X^0.

Proof: Let there be given $y \in X^0$ and a convergent sequence $x_n \to y$. We need to show that $u(x_n) \to u(y)$. We will consider two special cases: $(x_n) \subset X^1$ and $(x_n) \subset X^0$. If we show that for each of these the conclusion $u(x_n) \to u(y)$ holds, we are done, as any other sequence can be split into two subsequences, one in X^0 and the other in X^1, and each of these, if infinite, has to yield u values that converge to $u(y)$.

When we consider $(x_n) \subset X^1$ we are back to the first part of the proof, where we showed that $u(x_n)$ is convergent, and that its limit has to be $u(y)$.

Consider then a sequence $(x_n) \subset X^0$ such that $x_n \to y$ and assume that $u(x_n) \to u(y)$ doesn't hold. Then there exists $\varepsilon > 0$ such that, for infinitely many n's, $u(x_n) > u(y) + \varepsilon$, or that, for infinitely many n's, $u(x_n) < u(y) - \varepsilon$ (or both). For each n select a sequence $\left(x_n^k\right)_k \subset X^1$ such that $x_n^k \to_{k\to\infty} x_n$. For every m, pick n such that $\|x_n - y\| < \frac{1}{2m}$ and k such that $\left\|x_n^k - x_n\right\| < \frac{1}{2m}$ so that $(x_n^n) \subset X^1$ and $x_n^n \to_{n\to\infty} y$. However, $|u(x_n^n) - u(y)| \geq \varepsilon$, a contradiction. We thus conclude that u is continuous on X^0. $\square$

Next, we wish to show that the continuous extension we constructed represents $\succsim$ also on its extended domain, X^0. We do this in two steps. First, we observe the following:

Lemma 9 For all $x, y \in X^0$, if $u(x) > u(y)$ then $x \succ y$.

Proof: By definition of u, we can take sequences $(x_n), (y_n) \subset X^1$ such that $x_n \to x$ and $y_n \to y$. Letting $\varepsilon = u(x) - u(y) > 0$ choose N large enough so that for all $n \geq N$ we have $|u(x_n) - u(x)|, |u(y_n) - u(y)| < \varepsilon/3$. As u is continuous on X^1 we can also find $z^*, w^* \in X^1$ so that $u(z^*) = u(x) - \varepsilon/3$; $u(w^*) = u(y) + \varepsilon/3$. Thus $u(x_n) > u(z^*) > u(w^*) > u(y_n)$ for all $n \geq N$. A3 implies that $x \succ y$. $\square$

The next and final step of the proof is to show the converse, namely:

Lemma 10 For all $x, y \in X^0$, if $u(x) = u(y)$ then $x \sim y$.

Proof: We first prove an auxiliary claim:

Claim 1 *Assume that, for $z, w \in X^0$, $u(z) = u(w) = a$ but $z \succ w$. Let $(z_n), (w_n) \subset X^1$ converge to z and w respectively. Then $\exists N$ such that, $\forall n \geq N$ we have (i) $u(z_n) \geq a$ and (ii) $u(w_n) \leq a$.*

Proof of Claim: Suppose first that $u(z_n) < a$ occurs infinitely often. Let (n_k) be a sequence such that $u(z_{n_k}) < a$. Because $u(w_n) \to a$, for each such k we can find $m(n_k)$ such that $u\left(w_{m(n_k)}\right) > u(z_{n_k})$ and $m(n_k)$ increases in k. Thus we have two sequences $(z_{n_k}), \left(w_{m(n_k)}\right) \subset X^1$, converging to z and

w, respectively, with $w_{m(n_k)} \succ z_{n_k}$. By A2, we get $w \succsim z$, a contradiction. By a similar argument, if $u(w_n) > a$ occurs infinitely often, we select such a subsequence $u(w_{n_k}) > a$ and $u\left(z_{m(n_k)}\right) < u(w_{n_k})$ and $w \succsim z$ follows again. Thus, $\exists N$ such that, $\forall n \geq N$ we have both $u(z_n) \geq a$ and $u(w_n) \leq a$. $\square$

Equipped with this Claim we turn to prove the lemma. Assume that $x, y \in X^0$ satisfy $u(x) = u(y)$ but $x \succ y$. Because X^0 is connected and $\succsim$ satisfies A2, we have to have $z \in X^0$ such that $x \succ z \succ y$. Applying the same reasoning to z and y we can also get $w \in X^0$ such that $x \succ z \succ w \succ y$.

Let $a = u(x) = u(y)$. Applying Lemma 9, we know that $x \succ z \succ w \succ y$ and, indeed, $x \succsim z \succsim w \succsim y$ implies $u(x) \geq u(z) \geq u(w) \geq u(y)$ and thus we have $u(x) = u(z) = u(w) = u(y) = a$.

Let there be sequences $(x_n), (z_n), (w_n), (y_n) \subset X^1$ converging to x, z, w, y, respectively. Applying the Claim to $x \succ z$, we conclude that, from some N_1 on, $u(z_n) \leq a$. Applying the same Claim to $w \succ y$, we find that, from some N_2 on, $u(w_n) \geq a$. However, when we apply it to $z \succ w$ we find that, from some N_3 on, $u(z_n) \geq a$ and $u(w_n) \leq a$. For $n \geq \max(N_1, N_2, N_3)$ we have $u(z_n) = u(w_n) = a$. This means that $z_n \sim w_n$ and A2 yields $z \sim w$, a contradiction. $\square$

6.2 Examples

We use two continuity axioms, A2 and A3. A2 seems to be rather strong, and, as mentioned above, if we drop the comparability restriction, it is, per se,[15] stronger than the standard continuity assumption of consumer theory. Moreover, if we drop the comparability restriction, the two axioms are equivalent (for a weak order). Specifically, if we define

A2*. Universal Weak Preference Continuity: For all sequences $x_n \to x$ and $y_n \to y$, if $x_n \succsim y_n$ for all n, then $x \succsim y$.

A3*. Universal Strict Preference Continuity: For all sequences

[15] That is, without A1 necessarily assumed.

$x_n \to x$ and $y_n \to y$, and all $z, w \in X$, if $x_n \succsim z \succ w \succsim y_n$ for all n, then $x \succ y$.

We can state

Observation 1 *If $\succsim$ is a weak order on X, then A2* and A3* are equivalent.*

Proof: Assume first that $\succsim$ satisfies A2*. Then for the bundles in A3* we have $x \succsim z$ and $w \succsim y$, which implies $x \succ y$ by transitivity.

Next, assume that $\succsim$ satisfies A3*. We first claim that, for all sequences $x_n \to x$ and $y_n \to y$, if $x \succ y$, then there exists an N such that $x \succ y_n$ and $x_n \succ y$ for all $n > N$. To see this, suppose that the contrary holds. If $y_n \succsim x$ for infinitely many n's, then for these n's we have $y_n \succsim x \succ y \succsim y$, which by A3* implies $y \succ y$, a contradiction. Alternatively, $y \succsim x_n$ for infinitely many n's would imply $x \succsim x \succ y \succsim x_n$ and $x \succ x$.

To see that A2* holds, let there be given sequences $x_n \to x$ and $y_n \to y$, such that $x_n \succsim y_n$ for all n, and assume that, contrary to our claim, $y \succ x$. For all n large enough, $y \succ x_n \succsim y_n \succ x$ by the argument above. Fix such a k so that $y \succ x_k \succsim y_k \succ x$. Apply the argument again to conclude that, for some N, we have $y_n \succ y_k$ for all $n > N$. Since $x_n \succsim y_n$ for all n, we have by transitivity $x_n \succ y_k$ for all $n > N$. So we have $x_n \succ y_k \succ x \succsim x$ for all $n > N$, which by A3* implies $x \succ x$, an impossibility. $\square$

In light of this equivalence of the "universal" versions of the axioms (applying to all sequences, rather than only to comparable ones), one may wonder whether A3 is also needed, and, if so, maybe A3 can be assumed but A2 can be dispensed with. In the following we provide a few examples that show that none of the axioms is redundant. In the first five examples we have $n = 2$, $X = [0, 10]^2$ and $d = (1, 0)$, so that the principle is satisfied on the x_2 axis (X^0 consists of all the points with $x_1 = 0$) but not off the axis (X^0 consists of all the points with $x_1 > 0$). We define $\succsim$ by a numerical function v so that A1 is satisfied in all examples.

6.2.1 Example 1: A2 without A3 (I)

Let v be given by[16]:

$$v(x_1, x_2) = \begin{cases} 3 & x_1 = 0 \\ \sin\left(\frac{1}{x_1}\right) & x_1 > 0 \end{cases}$$

So the x_2 axis ($x_1 = 0$) is an indifference class that is preferred to anything else. Preference off the axis depend only on x_1, in a continuous way on the interior ($x_1 > 0$), but in a way that has no limit as we approach $x_1 = 0$.

To see that A2 is satisfied, consider $x_n \to x$ and $y_n \to y$ with $x_n \succsim y_n$ as in the antecedents of A2. Then if $x, y \in X^0$, the consequent $x \succsim y$ follows as $x \sim y$ for any $x, y \in X^0$. And if $x, y \in X^1$, then from some point on $x_n, y_n \in X^1$ and the consequent follows from the continuity of v on X^1. However, A3 isn't satisfied. More specifically, the claim of Lemma 4, which is an implication of A3, does not hold. To see this, define $x_n = \left(\frac{1}{\left(2n+\frac{1}{2}\right)\pi}, 1\right)$; $y_n = \left(\frac{1}{\left(2n+\frac{3}{2}\right)\pi}, 1\right)$ and $x = (0, 1)$ so that $x_n, y_n \to x$. Let $z = \left(\frac{2}{\pi}, 1\right)$ and $w = \left(\frac{2}{3\pi}, 1\right)$ so that $v(x_n) = v(z) = 1$ and $v(y_n) = v(w) = -1$. Thus, $x_n \succsim z$ and $w \succsim y_n$ but $w \succsim z$ doesn't hold. □

6.2.2 Example 2: A2 without A3 (II)

The previous example relies on the absence of a limit – preferences on X^1 have no "Cauchy sequences". The next example shows that this is only one problem that may arise, and that A3 may not hold even if preferences are very well-behaved on each of X^0, X^1. Let v be given by:

$$v(x_1, x_2) = \begin{cases} x_2 & x_1 = 0 \\ x_2 - 3 & x_1 > 0, x_2 < 5 \\ x_2 - 2 & x_1 > 0, x_2 = 5 \\ x_2 - 1 & x_1 > 0, x_2 > 5 \end{cases}$$

[16]Here and in the sequel we drop one set of parentheses for clarity. That is, $u_d((x_1, x_2))$ is denoted $u_d(x_1, x_2)$.

In the subspace $x_1 > 0$, $\succsim$ could also be represented by $v'(x_1, x_2) = x_2 - 2$ and it is clearly continuous there. But v is defined by taking $v'(x_1, x_2) = 3$ (corresponding to $x_2 = 5$) as a watershed, shifting the region $v'(x_1, x_2) > 3$ (corresponding to $x_2 > 5$) up by 1 and the region $v'(x_1, x_2) < 3$ (corresponding to $x_2 < 5$) down by 1. This generates "holes" in the range of u_d that could be skipped if we only had to worry about $x_1 > 0$. Yet, we cannot re-define u_d on this range to be continuous because we have points on the x_2 axis ($x_1 = 0$) that are in between preference-wise.

To see that A2 is satisfied, consider $x_n \to x$ and $y_n \to y$ with $x_n \succsim y_n$ as in the antecedents of A2. Then if $x, y \in X^0$, the consequent $x \succsim y$ follows because v is obviously continuous on X^0. And if $x, y \in X^1$, then from some point on $x_n, y_n \in X^1$ and the consequent follows from the fact that on X^1 the relation $\succsim$ could also be represented by v' which is continuous on X^1. However, A3 is violated. To see this, let $x_n = \left(1, 5 + \frac{1}{n}\right)$ and $y_n = \left(1, 5 - \frac{1}{n}\right)$ with $x = (1, 5)$ being their common limit. Take $z = (0, 4)$ and $w = (0, 3)$ so that $x_n \succsim z$ and $w \succsim y_n$ because $v\left(1, 5 + \frac{1}{n}\right) = 4 + \frac{1}{n} > v(0, 4)$ and $v(0, 3) = 3 > 2 + \frac{1}{n} = v\left(1, 5 - \frac{1}{n}\right)$. However, $w \succsim z$ doesn't hold. Thus, the claim of Lemma 4 is again violated. □

6.2.3 Example 3: Lemma 4

The next example satisfies the conclusion of Lemma 4 but not the other properties. Let v be defined by:

$$v(x_1, x_2) = \begin{cases} x_2 & x_1 = 0 \\ x_2 - 1 & x_1 > 0, x_2 < 5 \\ 9 - x_2 & x_1 > 0, x_2 \geq 5 \end{cases}$$

That is, as long as $x_2 \leq 5$ preferences are monotone in x_2 with a "jump" at the x_2 axis. However, when x_2 is above 5, the direction of preferences in the interior ($x_1 > 0$) reverses, but not on the axis.

These preferences do not satisfy A2. For example, let $x_n = \left(\frac{1}{n}, 4\right), y_n = \left(\frac{1}{n}, 6\right)$ with $x = (0, 4)$ and $y = (0, 6)$. Then we have $v(x_n) = v(y_n) = 3$ and thus $x_n \succsim y_n$, but $v(x) = 4 < 6 = v(y)$ so that $x \succsim y$ fails to hold.

At the same time, the conclusion of Lemma 4 holds. To see this, let $x_n \to x$ and $y_n \to x$. As v is uniformly continuous both on X^0 and on X^1, $\lim v(x_n)$ and $\lim v(y_n)$ exist and they are equal. This means that there can be no $a = v(z)$ and $b = v(w)$ such that $v(x_n) \geq a > b \geq v(y_n)$ for all n, and if $x_n \succsim z$ and $w \succsim y_n$ for all n, $w \succsim z$ has to follow.

Finally, these preferences also do not satisfy A3. To see this, we can take $x_n = \left(\frac{1}{n}, 4\right), y_n = \left(\frac{1}{n}, 7\right)$ so that $v(x_n) = 3$ and $v(y_n) = 2$. For $z = (0,3)$ and $w = (0,2)$ we have $v(z) = 3, v(w) = 2$ so that $x_n \succsim z \succ w \succsim y_n$. But the limit points, $x = (0,4)$ and $y = (0,7)$ do not satisfy $x \succ y$ (in fact, the converse holds, that is, $y \succ x$). $\square$

6.2.4 Example 4: A2 and Lemma 4 without A3

Next consider v defined by :

$$v(x_1, x_2) = \begin{cases} x_2 - 2 & x_1 > 0 \\ x_2 & x_1 = 0, \quad x_2 < 4 \\ 4 & x_1 = 0, \quad 4 \leq x_2 \leq 5 \\ x_2 - 1 & x_1 = 0, \quad x_2 > 5 \end{cases}$$

Thus, along the axis $x_1 = 0$, preferences are represented by a non-decreasing continuous function of x_2 that is constant on a given interval, and off it ($x_1 > 0$) they could also be represented by x_2.

We claim that these preferences satisfy A2 and the conclusion of Lemma 4 but not A3. Starting with A2, consider $x_n \to x$ and $y_n \to y$ with $x_n \succsim y_n$ as in the antecedents of A2. Then if $x_n, y_n \in X^0$, the consequent $x \succsim y$ follows because v is continuous on X^0. And if $x_n, y_n, x, y \in X^1$, the consequent follows from the fact that on X^1 the relation $\succsim$ could also be represented by $v' = x_2$. We are left with the interesting case in which $x_n, y_n \in X^1$ but $x, y \in X^0$. Because $x_n \succsim y_n$, we know that the second component of x_n is at least as high as is that of y_n, and it follows that the same inequality holds in the limit and $x \succsim y$.

The conclusion of Lemma 4 also holds because v is uniformly continuous

on each of X^0 and X^1. Thus, $x_n \to x$ and $y_n \to x$ imply that $\lim v(x_n) = \lim v(y_n)$ (and that both exist).

However, A3 fails to hold. To see this, consider $x_n = \left(\frac{1}{n}, 4\right), y_n = \left(\frac{1}{n}, 5\right)$ with $x = (0,4)$ and $y = (0,5)$. For $z = (0,3)$ and $w = (0,2)$ we have $u_d(z) = 3, u_d(w) = 2$ so that $y_n \succsim z \succ w \succsim x_n$. But for limit points $x \sim y$, in violation of the axiom. □

6.2.5 Example 5: A3 without A2

Finally, we show that A3 does not imply A2. Let

$$v(x_1, x_2) = \begin{cases} -1 & x_1 > 0 \\ x_2 & x_1 = 0 \end{cases}$$

That is, the entire X^1 is a single indifference class that is below, preference-wise, the entire x_2 axis. We claim that these preferences satisfy A3 but not A2.

To see that A3 holds, consider $(x_n), (y_n)$ and x, y, z, w in X such that $x_n \to x$ and $y_n \to y$ and $x_n \succsim z \succ w \succsim y_n$. If $(x_n), (y_n) \subset X^0$ then we have $x, y \in X^0$. Because v is simply x_2 on X^0, the conclusion follows. If $(x_n), (y_n) \subset X^1$ we cannot have $x_n \succsim z \succ w \succsim y_n$ because $x_n \sim y_n$. Thus, A3 holds.

However, A2 can easily seen to be violated. For example, $x_n = \left(\frac{1}{n}, 4\right), y_n = \left(\frac{1}{n}, 5\right)$ satisfy $x_n \succsim y_n$ but at the limit we get $(0,5) \succ (0,4)$. □

6.2.6 Example 6: The Role of Connectedness

The following example shows that for Theorem 2 to hold, the set X has to be connected. Let

$$X = \left\{ (x_1, x_2) \;\middle|\; \begin{array}{c} 0 \le x_1 \le 1 \\ 0 \le x_2 \le 1 \\ \textit{or} \\ 2 \le x_2 \le 3 \end{array} \right\}$$

and define the following two functions on X:

$$u(x_1, x_2) = \begin{cases} -x_1 & 0 \leq x_2 \leq 1 \\ x_1 & 2 \leq x_2 \leq 3 \end{cases}$$

$$v(x_1, x_2) = \begin{cases} -x_1 & 0 \leq x_2 \leq 1 \\ x_1 + 1 & 2 \leq x_2 \leq 3 \end{cases}$$

Define $\succsim$ on X by maximization of v. As v is continuous, $\succsim$ satisfies axioms A1-A3. Note that u, restricted to $X^1 = \{ (x_1, x_2) \mid x_1 > 0 \}$, represents $\succsim$ as well. Indeed, it has a continuous extension to X – u itself. However, it does not represent $\succsim$ on X^0, as u is constant on X^0 which isn't an equivalence class of $\succsim$ (say, $(2,0) \succ (1,0)$). □

7 Appendix B: Proofs of Representation Results

It will be convenient to introduce the following definition of a binary relation P on X^0:

Definition 2 *For $x, y \in X^0$, we say that xPy if there exists $z \in X^0$ and a sequence $z_n \to z$ with $(z_n) \subset X^1$ such that $x \succsim z$ and $z_n \succsim y$.*

Observe that, if we had no discontinuity between X^0 and X^1, the relation P could be expected to be equal to $\succsim$: if xPy, the conditions $z_n \to z$ and $z_n \succsim y$ would simply imply that $z \succsim y$, and $x \succsim y$ would follow by transitivity. Conversely, if $x \succsim y$, one could expect an open neighborhood of x to contain points z_n such that $z_n \succsim y$ even though $z_n \in X^1$ (for example, monotonicity would insure that this is the case). However, in the presence of discontinuity between X^1 and X^0, this is no longer the case. As explained above in the context of A4, we should expect z to be strictly better than y; indeed, intuitively, "z should be better than y at least by the cost of the principle". And the same should hold for any $x \in X^0$ such that $x \succsim z$.

Using this definition, the Archimedean axiom can be written as follows.

A4 Archimedeanity (in P terms): Let $(x_n) \subset X^0$ be such that $x_{n+1} P x_n$ ($x_n P x_{n+1}$) for all $n \geq 1$. Then there does not exist $\hat{x} \in X$ such that $\hat{x} \succsim x_n$ ($x_n \succsim \hat{x}$) for all $n \geq 1$.

This new formulation of A4 is simply a re-statement of the axiom in terms of the relation P. We therefore do not re-name the axiom.

7.1 Proof of Theorem 1

Necessity

We first note that, if a representation as in (7) exists, A1-A5 hold. A1 is immediate as $\succsim$ is represented by a numerical function, and A5 – because it is non-constant.

A2 holds because u is continuous over all of X: let there be given comparable sequences $x_n \to x$ and $y_n \to y$ with $x_n \succsim y_n$. If they are comparable because of Condition A, then for some i we have $(x_n) \subset X^i$ and $x \in X^i$. Whether $i = 0$ or $i = 1$ we have $u_d(x_n) \to u_d(x)$, and the same holds for the sequence $y_n \to y$ (which is in one of the X^j as is its limit). If, however, the sequences are comparable because of Condition B, the condition $x_n \succsim y_n$ implies that $u_d(x_n) \geq u_d(y_n)$ and this implies $u(x_n) \geq u(y_n)$ whether $(x_n), (y_n) \subset X^0$ (and then $u(x_n) = u_d(x_n)$; $u(y_n) = u_d(y_n)$) or $(x_n), (y_n) \subset X^1$ (and then $u(x_n) = u_d(x_n) + \gamma$; $u(y_n) = u_d(y_n) + \gamma$). This means that $u(x) \geq u(y)$. It follows that $x \succsim y$ whether $x, y \in X^0$ (and then $u_d(x) = u(x)$; $u_d(y) = u(y)$) or $x, y \in X^1$ (and then $u_d(x) = u(x) - \gamma$; $u_d(y) = u(y) - \gamma$).

We now turn to prove the necessity of A3. Let there be given two comparable sequences, $x_n \to x$ and $y_n \to y$, and $z, w \in X$ such that $x_n \succsim z \succ w \succsim y_n$. Then, we have that $u_d(x_n) - u_d(y_n) \geq u_d(z) - u_d(w) > 0$. Furthermore, we note that $u_d(x_n) - u_d(y_n) \to u_d(x) - u_d(y)$ because $x_n \to x$ and $y_n \to y$ are comparable (if it is because of Condition A, then $u_d(x_n) \to u_d(x)$ and

$u_d(y_n) \to u_d(y)$, and if because of Condition B, $u_d(x_n) \to u_d(x) - \gamma$ and $u_d(y_n) \to u_d(y) - \gamma$). Thus, $x \succ y$.

To prove that A4 is also necessary for the representation, observe that, for $x, y \in X^0$, the relation xPy implies that $u(x) - u(y) \geq \gamma$. This is true because, if $(z_n) \subset X^1$ is such that $z_n \to z$, $x \succsim z$, and $z_n \succsim y$, we have $u_d(z_n) \geq u_d(y)$ or $u(z_n) - \gamma \geq u(y)$. As $z_n \to z$ and u is continuous, $u(z) - u(y) \geq \gamma$ follows. Finally, as $x \succsim z$ we also have $u_d(x) \geq u_d(z)$ and $u(x) \geq u(z)$ (as $x, z \in X^0$ imply $u_d(x) = u(x); u(z) = u_d(z)$) and $u(x) - u(y) \geq \gamma$. Hence an infinite increasing (decreasing) P chain will have u values that are unbounded from above (below), an impossibility.

Sufficiency

We now turn to prove the sufficiency of the axioms. We remind the reader of Lemmas 1-4 from Appendix A, which will be used here. In particular, we recall that $\succsim$ is continuous on X^1, and thus there exists a continuous function v that represents $\succsim$ on X^1. We can assume w.l.o.g. that v is bounded, and then use Theorem 2 to extend v continuously to all of X so that it represents $\succsim$ on X^0 as well.

Proof overview The strategy of the proof is as follows. We will construct a continuous function u_d on X^0 that represents $\succsim$ and that also represents P by γ differences (that is, for any $x, y \in X^0$, xPy if and only if $u_d(x) - u_d(y) \geq \gamma$). To this end we can start out with any continuous function that represents $\succsim$ on those $x \in X^0$ for which there are no $y \in X^0$ such that xPy. It will be convenient to use the function $v(\cdot) + \gamma$ on that set, and to extend it to the rest of X^0 while respecting the representation of P by γ differences. Intuitively, this can be done because the utility function is ordinal, and we can choose various ways to pin down one such function. The way we do it for this theorem is by setting "stepping off X^0 into X^1" to be a constant measuring rod. If γ is, intuitively, the cost of giving up the principle, we choose a utility

function that measures utility differences in terms of the "number of times the principle has been compromised".

Any element of X^1 that has a $\succsim$-equivalent in X^0 will have to have the same u_d value, and we will show that the resulting function is continuous on X^1 as well. Moreover, we will show that the function so constructed has a constant "jump" of γ between any sequence in X^1 that converges to a limit in X^0. That is, if we subtract γ from u_d on X^0, but not on X^1, we obtain a continuous function ($u(\cdot) - \gamma$ in the representation).

For elements of X^1 which are strictly better or strictly worse than all elements of X^0 will have to be assigned a utility value that represents $\succsim$, and that also behaves in this very specific way near the boundary, X^0, namely, having a "jump" of γ at the limit. Strictly better elements will not pose a problem, because, as will be clear from the proof, they cannot converge to an element in X^0. (Intuitively, the reason is that there is a preference *for* the principle.) By contrast, strictly worse elements might converge to a limit point in X^0 (and some such sequences will be bound to exist). However, these limit points $x \in X^0$ will be in the "lower-preference" part of X^0 where there are no $y \in X^0$ such that xPy. This means that the choice of u_d in this set was $v(\cdot) + \gamma$ (where v is the continuous extension) and the constant γ "jump" at these points holds as well.

We now turn to the proof of the sufficiency part.

Proof. As explained above, we would like to think of xPy as saying "x is better than y, to an extent that justifies giving up the principle, that is, even more than stepping off X^0 into X^1". Thus, the relation xPy should mean that the utility of x is higher than that of y, by at least as much as a gap that is the cost of the principle. To support this interpretation we would like first to see that xPy implies that x is strictly preferred to y:

Lemma 11 For $x, y \in X^0$, if xPy then $x \succ y$.

Proof: Assume not. In this case, there is a sequence $z_n \to z$, $(z_n) \subset X^1$,

$x \succsim z$, and $z_n \succsim y$ but $y \succsim x$. By transitivity of $\succsim$, we also get $z_n \succsim x$. Then, by definition of P (with the same sequence $z_n \to z$), we have xPx. This means that we can define a sequence $x_n = x \in X^0$ such that $x_{n+1} P x_n$ for all n and the sequence is bounded (by x itself $\hat{x} \equiv x \succsim x_n = x$), in violation of A4. $\square$

Next, we would expect the relation P to behave "monotonically" relative to $\succsim$: if xPy, then the relation should hold if we make x better or y worse (or both):

Lemma 12 For $x, y, w \in X^0$, if xPy then (i) $y \succsim w$ implies xPw, and (ii) $w \succsim x$ implies wPy.

Proof: Suppose that $x, y, z \in X^0$ and $(z_n) \subset X^1$ are given, such that $z_n \to z$, $x \succsim z$ and $z_n \succsim y$. In case (i), $z_n \succsim y \succsim w$ and by transitivity $z_n \succsim w$, which implies xPw by definition of the relation P. As for (ii), $w \succsim x$ and $x \succsim z$ imply $w \succsim z$ and the definition of P yields wPy. $\square$

We can now use a more convenient definition of the relation P: to say that xPy, we can use a convergent sequence that is equivalent to (rather than at least as preferred as) y. Explicitly,

Lemma 13 For $x, y \in X^0$, if xPy, then there exists $z \in X^0$ and a sequence (z_n) with $z_n \in X^1$ such that $z_n \to z$, $x \succsim z$ and $z_n \sim y$.

Proof. Assume that $x, y \in X^0$ satisfy xPy, and that $z \in X^0$ and (z_n) with $z_n \in X^1$ satisfy $z_n \to z$, $x \succsim z$ and $z_n \succsim y$. We argue that, for each n, there exists $\alpha_n \in (0, 1]$ such that $w_n \equiv \alpha_n z_n + (1 - \alpha_n)\, y \in X^1$ satisfies $w_n \sim y$. To see this, first notice that, if $z_n \sim y$ we can set $\alpha_n = 1$ and we are done. Assume, then, that $z_n \succ y$. If there exists $\beta \in (0, 1]$ such that $y \succ \beta z_n + (1 - \beta)\, y$ then we have $z_n \succ y \succ \beta z_n + (1 - \beta)\, y$, with $z_n, \beta z_n + (1 - \beta)\, y \in X^1$, and we can apply Lemma 2 to obtain the existence of a point on the interval $[\beta z_n + (1 - \beta)\, y, z_n]$ that satisfies indifference to

y, and that point is evidently also on the interval $[y, z_n]$ and we are done. However, if such a β does not exist, we have $\beta z_n + (1-\beta)\, y \succ y$ for all $\beta > 0$. Taking a subsequence $\beta_k \searrow 0$, with $\beta_k z_n + (1-\beta_k)\, y \to y$, we obtain yPy, in contradiction to Lemma 11.

We conclude that there are $\alpha_n \in (0,1]$ such that $w_n \equiv \alpha_n z_n + (1-\alpha_n)\, y \sim y$; we observe that $w_n \in X^1$ holds because $\alpha_n > 0$. Choose a convergent subsequence of α_n, say $\alpha_{n_k} \to \alpha^*$. Then $w_{n_k} \to w^* \equiv \alpha^* z + (1-\alpha^*)\, y \in X^0$. We need to show that $x \succsim w^*$. To see this, observe that $z_{n_k} \succsim w_{n_k}$ (because $z_{n_k} \succsim y$ and $w_{n_k} \sim y$), $z_{n_k} \to z$,$w_{n_k} \to w^*$, while $(z_{n_k})_k, (w_{n_k})_k \subset X^1$ and $z, w^* \in X^0$. Hence $(z_{n_k})_k \to z$ and $(w_{n_k})_k \to w^*$ are comparable and we can use A2 to conclude that $z \succsim w^*$ and $x \succsim w^*$ follows by transitivity. □

Because $\succsim$ satisfies the standard assumptions of consumer theory on each of X^0, X^1, we know that it can be represented by a continuous function on each of them separately. The strategy of the proof is to choose a continuous representation on X^0, take a monotone and continuous transformation thereof to obtain another representation, u and a number $\gamma > 0$ such that the relation P would be roughly equivalent to a difference of at least γ in the level of u. More precisely, define

$$X_P^0 = \left\{ y \in X^0 \;\middle|\; \exists x \in X^0,\, xPy \right\}$$

and for $x \in X^0, y \in X_P^0$ we will have

$$xPy \qquad \Leftrightarrow \qquad u(x) - u(y) \geq \gamma > 0$$

We will then extend the function u to X^1. To this end, it will be useful to know some facts about continuous representations of $\succsim$ on X^0.

Lemma 14 Let there be given a continuous function $u : X^0 \to \mathbb{R}$ that represents $\succsim$ (on X^0). Let $y \in X_P^0$. Then there exists $\gamma(y) > 0$ such that, for every $x \in X^0$,

$$xPy \qquad \Leftrightarrow \qquad u(x) - u(y) \geq \gamma(y) \tag{8}$$

Furthermore, $\gamma(y)$ can be extended to all of X^0 so that $w \succsim y$ iff $u(w) + \gamma(w) \geq u(y) + \gamma(y)$ (for all $y, w \in X^0$).

Proof: Define $P_{y+} = \{ x \in X^0 \mid xPy \}$.

Case (a): Let us first consider $y \in X_P^0$ so that $P_{y+} \neq \varnothing$. Consider

$$u(P_{y+}) = \left\{ u(x) \in u(X^0) \mid xPy \right\}$$

By Lemma 11, $u(y) < a$ for all $a \in u(P_{y+})$. By Lemma 12, $u(P_{y+})$ is an interval.

We wish to show that it contains its infimum. Let $a = \inf u(P_{y+})$. For $k \geq 1$, let $x^k \in X^0$ be such that $a \leq u\left(x^k\right) < a + \frac{1}{k}$. Because $x^k P y$, by Lemma 13, there exist (i) $z^k \in X^0$ and (ii) $\left(z_n^k\right)_{n \geq 1}$ with $z_n^k \in X^1$ such that $z_n^k \to z^k$, $x^k \succsim z^k$ and $z_n^k \sim y$. This implies that, for every k, l, m, n, $z_n^k \sim z_m^l (\sim y)$. Because $\left(z_n^k\right), \left(z_m^l\right) \subset X^1$ converge to $z^k, z^l \in X^0$ respectively, we also get, by A2, $z^k \sim z^l$. We know that u represents $\succsim$ on X^0 and this means that $u\left(z^k\right) = u\left(z^l\right)$ and this holds for all k, l only if $u\left(z^k\right) = a$. Thus, $a = \min u(P_{y+})$ and $a > u(y)$. It remains to define $\gamma(y) = a - u(y) > 0$. Clearly, for every y such that $P_{y+} \neq \varnothing$, $\gamma(y)$ is bounded from above (by $u(x) - u(y)$ for any $x \in P_{y+}$).

Observe that $\gamma(y)$ is uniquely defined for all $y \in X_P^0$. We turn to prove that for this function γ, $u + \gamma$ also represents $\succsim$ for alternatives y, w in this range.

In the construction above, $u(y) + \gamma(y) = \min u(P_{y+})$. If $w \succsim y$, Lemma 12 implies that $P_{w+} \subset P_{y+}$ and thus $\min u(P_{w+}) \geq \min u(P_{y+})$, so that $u(w) + \gamma(w) \geq u(y) + \gamma(y)$ follows. We wish to show that the inequality is strict if $w \succ y$. Let $x \in X^0$ be a $\succsim$-minimal element in P_{w+}, that is, $u(x) = u(w) + \gamma(w)$. We wish to show that there exists x' with $u(x') < u(x)$ such that $x'Py$ still holds (while $x'Pw$ doesn't). Because xPw, by Lemma 13 there exists $z \in X^0$ and a sequence $(z_n)_{n \geq 1}$ with $z_n \in X^1$ such that $z_n \to z$, $x \sim z$ and $z_n \sim w$. (Note that $x \sim z$ follows from the minimality of x.) Hence, $z_n \succ y$. We proceed in a way that mimics the proof of Lemma 13:

for each z_n we can find $\alpha_n \in (0,1]$ such that $t_n \equiv \alpha_n z_n + (1-\alpha_n)\, y \in X^1$ satisfies $t_n \sim y$ (or else yPy would follow).

Taking a convergent subsequence of α_n, say $\alpha_{n_k} \to \alpha^*$, we have $t_{n_k} \to t^* \equiv \alpha^* z + (1-\alpha^*)\, y \in X^0$. We thus have two sequences $(z_{n_k}), (t_{n_k}) \subset X^1$, with $z_{n_k} \sim w \succ y \sim t_{n_k}$ and $z_{n_k} \to z, t_{n_k} \to t^*$ with $z, t^* \in X^0$. Observe that $(z_{n_k}) \to z, (t_{n_k}) \to t^*$ are comparable. Hence A3 implies that $z \succ t^*$. Thus we can find $x' \in X^0$ with $u(x') \in (u(t^*), u(z))$. As z (and x) was selected to have the lowest possible u in $u(P_{w+})$, $x'Pw$ doesn't hold, while $x'Py$ does.

Case (b): For $y \in X^0 \backslash X_P^0$ we set $\gamma(y)$ to be a constant, defined as follows. Let $\bar{u} = \sup_{z \in X_P^0} u(z)$. This sup may or may not be a max.[17] Define $\gamma(y) = \lim_{n\to\infty} \sup \left\{ \gamma(z) \,\middle|\, \bar{u} - \frac{1}{n} < u(z) \leq \bar{u} \right\}$. It is finite because $\gamma(y)$ is bounded from above by $(u(x) - \bar{u} + 1)$ for any $x \in P_{z+}$. Because $\gamma(y)$ is constant for all $y \in X^0 \backslash X_P^0$, and because u represents $\succsim$ for alternatives y, w in this range, so does $u + \gamma$. Next, observe that $\sup_{z \in X_P^0} [u(z) + \gamma(z)] = \sup_{z \in X^0} u(z)$ and, for $y \in X^0 \backslash X_P^0$ and $w \in X_P^0$ we have

$$\begin{aligned} u(y) + \gamma(y) &\geq \sup_{z \in X^0} u(z) \geq u(w) + \gamma(w) \\ u(y) + \gamma(y) &> u(w) + \gamma(w) \end{aligned}$$

that is, the value $\sup_{z \in X^0} u(z)$ might be obtained by $u(\cdot) + \gamma(\cdot)$ on X^0 or on $X^0 \backslash X_P^0$ but not on both, so that $u + \gamma$ represents $\succsim$ on the entire range. □

Lemma 15 Let there be given a continuous function $u : X^0 \to \mathbb{R}$ that represents $\succsim$ (on X^0). There exists a continuous function $\phi : u(X^0) \to \mathbb{R}$ such that, for every $x \in X^0, y \in X_P^0$

$$xPy \qquad \Leftrightarrow \qquad u(x) - u(y) \geq \phi(u(y))$$

and $u(\cdot) + \phi(u(\cdot))$ also represents $\succsim$ on X^0.

[17] For example, for $n = 2$, $X = [0,10]^2$ and $d = (1,0)$ consider $u_1(x_1, x_2) = x_2 + x_1$ and $u_2(x_1, x_2) = x_2 + (x_1 - 1)^2$. In both cases define the relation by the function u_i and $\gamma = 1$. In the case of u_1 the relation P is a closed subset of $X^0 \times X^0$ an $\bar{u} = 9$ is the max of $u(z)$ over X_P^0, whereas for u_2 P isn't closed, and the point $(9,0)$ is not in X_P^0, leaving $\bar{u} = 9$ the sup of the utility in X_P^0.

Proof: Use Lemma 14 to define $\gamma : X^0 \to \mathbb{R}$ such that $u(\cdot) + \gamma(\cdot)$ represents $\succsim$ on X^0 and xPy iff $u(x) - u(y) \geq \gamma(y)$ whenever $y \in X_P^0$ as above. Observe that, for any $y, w \in X^0$, we have $w \succsim y$ iff $u(w) + \gamma(w) \geq u(y) + \gamma(y)$. Hence $w \sim y$ implies $u(w) + \gamma(w) = u(y) + \gamma(y)$ and, since $u(w) = u(y)$ also holds in this case, we also get $\gamma(w) = \gamma(y)$. It follows that we can define a function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\gamma(y) = \phi(u(y))$.

Observe that ϕ is uniquely defined for all values $\bar{u} = u(y)$ such that $y \in X_P^0$ (a condition that doesn't depend on the specific choice of $y \in u^{-1}(\bar{u})$). We now wish to show that ϕ is continuous on that range (recalling that it is constant, hence continuous, on the range of higher u values). Let there be given $\bar{u} \in range(u)$ and $\left(u^k\right)_{k\geq 1}$ so that $u^k \in range(u)$ and $u^k \to \bar{u}$ (as $k \to \infty$) and we need to show that $\phi\left(u^k\right) \to \phi(\bar{u})$. Assume not. Then there exists $\varepsilon > 0$ such that (i) there are infinitely many k's for which $\phi\left(u^k\right) < \phi(\bar{u}) - \varepsilon$ or (ii) there are infinitely many k's for which $\phi\left(u^k\right) > \phi(\bar{u}) + \varepsilon$.

In case (i), let $y \in u^{-1}(\bar{u})$ and $y^k \in u^{-1}\left(u^k\right)$ for k from some k_0 on (obviously, with $y \in X_P^0$ and $y^k \in X_P^0$ for all k). As u is continuous, we can also choose such a y and a corresponding sequence so that $y^k \to y$. Let $t, t' \in X^0$ be such that $u(y) + \phi(u(y)) = u(t) > u(t') > u\left(y^k\right) + \phi\left(u\left(y^k\right)\right)$ for all $k \geq k_0$, so that $t \succ t'$, tPy^k,$t'Py^k$ for all k, tPy but not $t'Py$. As tPy we can select a sequence $(z_n) \subset X^1$ with $z_n \to z \in X^0$, $t \succsim z$ and $z_n \sim y$. By the choice of t (as a u-minimal element such that tPy), $u(t) = u(z)$. As $t'Py^k$, there is, for each k, a sequence $\left(w_n^k\right) \subset X^1$ such that $w_n^k \to w^k \in X^0$, $t' \succsim w^k$ and $w_n^k \sim y^k$. As above, select a convergent subsequence of the diagonal to get a sequence $(w_n^n) \subset X^1$ such that $w_n^n \to w \in X^0$, $t' \succsim w \in X^0$ and $w_n^n \sim y^n$. By transitivity, $z \sim t \succ t' \succsim w$. Observe that $z_n \to z$ and $w_n^n \to w$ are comparable, and we also have $z \succ w$. We now use Lemma 5 for $y_n = y^n \to y$ and $x_n = x = y$. Clearly, x_n, y_n, x, y are all in X^0, and thus $y_n \to y$ and $x_n \to y$ are also comparable. Lemma 5 implies $y \succ y$, a contradiction.

In case (ii) select $t, t' \in X^0$ be such that $u(y) + \phi(u(y)) = u(t) <$

$u(t') < u(y^k) + \phi(u(y^k))$ for all $k \geq k_0$, so that $t' \succ t$, tPy and $t'Py$ hold, but tPy^k,$t'Py^k$ do not hold for any k. For each k, pick t^k such that $u(t^k) = u(y^k) + \phi(u(y^k))$, that is, a u-minimal element such that $t^k P y^k$. Let $(z_n^k) \subset X^1$ be such that $z_n^k \to z^k \in X^0$, $t^k \succsim z^k$ and $z_n^k \sim y^k$. Let $(z_n) \subset X^1$ be such that $z_n \to z \in X^0$, $t \succsim z$ and $z_n \sim y$. By the choice of t,(t^k) as minimal elements, $t \sim z$ and $t^k \sim z^k$. Select a convergent subsequence of $z_k^k \to z^* \in X^0$. Because $z^k \succsim t'$ (and $z^k \in X^0$) we have $z^* \succsim t' \succ t$. The contradiction follows from Lemma 5 as in the previous case.

We now turn to complete the proof. Recall that, by Theorem 2, there exists a continuous function $v : X \to \mathbb{R}$ that represents $\succsim$ on X^0 and on X^1. Without loss of generality, we can assume that v is bounded. By A5, it isn't constant on X^0. We can therefore further assume, without loss of generality, that $\inf_{x \in X^0} v(x) = 0$ and $\sup_{x \in X^0} v(x) = 1$. Let $b = \inf_{x \in X} v(x)$ and $a = \sup_{x \in X} v(x)$ so that $b \leq 0 \leq 1 \leq a$.

Next, we define a continuous $u : X \to \mathbb{R}$ and $\gamma > 0$ such that $u_d(x) = u(x) - \gamma \mathbf{1}_{\{x \in X^1\}}$ represents $\succsim$. We first define $u_d = u$ on X^0, and $\Delta > 0$ such that u represents $\succsim$ on X^0, and (u, Δ) jointly represent P on X^0 by

$$xPy \qquad \Leftrightarrow \qquad u(x) - u(y) \geq \Delta \tag{9}$$

We then proceed to define u on X^1 and γ.

Step 1: Definition of $u_d = u$ on X^0

If $P = \varnothing$ define $u = v$ and $\Delta = 2$. Clearly, (9) holds.

Otherwise, if $P \neq \varnothing$, we construct a partition of X^0 into countably many subsets X_k^0 for $k \in \mathbb{Z}$ in such a way that, if $x \in X_k^0$ and $y \in X_l^0$, then $k > l + 1$ implies xPy and $k \leq l$ implies that xPy does not hold. In fact, if $y \in X_l^0$, the least preferred x such that xPy (if such exists) will have to lie in X_{l+1}^0. In the construction we describe below, it is possible that finitely many subsets suffice, and the rest can be empty. Each X_k^0 will correspond to a half-open interval of v values (apart from one, possibly). There are uncountably many partitions that satisfy these conditions, but, once we decide on a given element of a partition, the rest are uniquely determined.

First, we define a function

$$S : X^0 \times X^0 \to \mathbb{Z}$$

to measure, roughly, the longest P-chain between two elements in X^0. Specifically, for $x, y \in X^0$, if $x \succsim y$, define

$$S(y,x) = \max \left\{ k \in \mathbb{Z} \;\middle|\; \begin{array}{c} k \geq 0 \\ \exists z_{0,\ldots,} z_k \in X^0 \\ z_0 = x \quad z_k = y \\ z_i P z_{i+1} \quad \forall i \leq k-1 \end{array} \right\}$$

Thus, $S(y,x) = 0$ for all $y \sim x$, and $x' \succsim x \succsim y \succsim y'$ implies $S(y',x') \geq S(y,x)$. For all $x \succsim y$, $S(y,x) < \infty$ by A4. Also, it is easy to see that, for $x \succsim y \succsim z$, we have

$$S(z,y) + S(y,x) \leq S(z,x) \leq S(z,y) + S(y,x) + 1$$

Next, for $x, y \in X^0$ with $y \succ x$, set

$$S(y,x) = -S(x,y) - 1$$

so that $S(y,x) + S(x,y) = -1$ for all $x \nsim y$.

We finally define u on X^0. Distinguish between two cases:

Case 1a: For every $x \in X^0$ there exists $y \in X^0$ such that xPy.

In this case, where there are infinitely decreasing P chains, should (9) hold, u should be unbounded from below. We can use almost any starting point to construct u – it is convenient to start with a point that isn't a maximal element of P. More explicitly, select an $x_0 \in X^0$ with $P_{x_0+} \neq \varnothing$. For any integer k, let $X_k^0 = \{ y \in X^0 \mid S(x_0, y) = k \}$. For all $y \in X_0^0$ (that is, $y \succsim x_0$ but not yPx_0), set $u(y) = v(y) - v(x_0)$ (in particular, $u(x_0) = 0$).

Let $\Delta = \sup_{X_0^0} u(y)$. Note that, by Lemma 14, $\Delta > 0$. Assuming that u is defined for all $y \in X_k^0$ for $k \geq 0$, extend it to X_{k+1}^0 as follows: for each $x \in X_{k+1}^0$ there exists $y \in X_k^0$ such that $v(x) = v(y) + \phi(v(y))$ where ϕ is the function constructed in Lemma 15 for v (and by Lemma 15,

this is the highest y that satisfies xPy). Set $u(x) = u(y) + \Delta$. Similarly, assuming that u is defined for all $y \in X_k^0$ for $k \leq 0$, extend it to X_{k-1}^0 by $u(x) = u(y) - \Delta$ for $x \in X_{k-1}^0$ and $y \in X_k^0$ such that $v(y) = v(x) + \phi(v(x))$. It is straightforward to verify that u so constructed is a continuous strictly monotone transformation of v and thus represents $\succsim$ on X^0. Consistent with our notation, we also define $u_d = u$ on X^0.

Case 1b: There exists $x \in X^0$ such that, for no $y \in X^0$ do we have xPy.

If there exists a v- (equivalently, a $\succsim$-) minimal element in X^0, we denote it by x_0 and proceed as in Case 1a. If not, let

$$\alpha = \sup \left\{ v(x) \;\middle|\; \begin{array}{c} x \in X^0 \\ \nexists y \in X^0, \quad xPy \end{array} \right\}$$

so that $v(x) > \alpha$ implies that xPy for some $y \in X^0$ and $v(x) < \alpha$ implies that xPy for no $y \in X^0$. It has to be the case that $\alpha > 0$ because $\alpha = 0$, in the absence of a minimal element, would mean that we are in Case 1a (where each $x \in X^0$ P-dominates at least one other element).

Define $u(x) = v(x)$ for all x with $v(x) \leq \alpha$ and $\Delta = \alpha$. For x with $v(x) > \alpha$ we repeat the construction above, with X_k^0 including all elements $x \in X^0$ for which the maximal decreasing P-chain is of length k.

Step 2: Definition of u on X^1 and of γ

To extend the function to all of X, partition X^1 into three sets,

$$\begin{aligned} X^{1\prec} &= \left\{ x \in X^1 \;\middle|\; \begin{array}{c} \forall y \in X^0 \\ x \prec y \end{array} \right\} \\ X^{1\sim} &= \left\{ x \in X^1 \;\middle|\; \begin{array}{c} \exists y \in X^0 \\ x \sim y \end{array} \right\} \quad X^{1\succ} = \left\{ x \in X^1 \;\middle|\; \begin{array}{c} \forall y \in X^0 \\ x \succ y \end{array} \right\} \end{aligned}$$

If $X^{1\sim} \neq \varnothing$, each of $X^{1\prec}$,$X^{1\succ}$ may be empty or not. However, if $X^{1\sim} = \varnothing$ we have to have $X^{1\prec} \neq \varnothing$: otherwise ($X^{1\succ} = X^1$) all $x \in X^1$ and $y \in X^0$ will satisfy $x \succ y$ and yPy would follow. Further, in this case, since $X^{1\sim} = \varnothing$ and $X^{1\prec} \neq \varnothing$ we also have $X^{1\succ} = \varnothing$, by Lemma 2. We will therefore split the definition according to the emptiness of $X^{1\sim}$.

Case 2a: $X^{1\sim} = \varnothing$

In this case we have $X^{1\sim} = X^{1\succ} = \varnothing$ as well as $P = \varnothing$ (as no element in X^1 is ranked as high as any in X^0). Define $u(x) = v(x)$ for all $x \in X^1 = X^{1\prec}$, and set $\gamma = 2(a-b) \geq \Delta$. On X^1, $u_d(x) = v(x) - \gamma$. Thus $u = v$ is a continuous function on all of X, u_d represents $\succsim$ on X^0 as well as on X^1, and it also satisfies $u_d(x) < u_d(y)$ for every $x \in X^1$ and every $y \in X^0$.

Case 2b: $X^{1\sim} \neq \varnothing$

It is simpler to first define u_d that would represent $\succsim$ on the entire space, and then find the $\gamma > 0$ such that $u(x) = u_d(x) + \gamma \mathbf{1}_{\{x \in X^1\}}$ is continuous.

We define u_d on $X^{1\sim}$ in the obvious way (in order to represent $\succsim$): for $x \in X^{1\sim}$, let $y \in X^0$ be such that $x \sim y$ and define $u_d(x) = u_d(y)$. This function represents $\succsim$ on $X^0 \cup X^{1\sim}$. We wish to show that it is continuous on $X^{1\sim}$.

Claim: $u_d : X^0 \cup X^{1\sim} \to \mathbb{R}$ is continuous (also) on $X^{1\sim}$.

Proof of the Claim: Let there be $(x_n) \to x$ in $X^{1\sim}$ and select corresponding $(y_n), y$ in X^0 (so that $x \sim y$ and $x_n \sim y_n$). Assume first that $x_1 \succ x$ and that $x_1 \succsim x_n \succsim x$ for all n. We wish to show that under these conditions $(x_n) \to x$. A symmetric argument would apply to the case $x_1 \prec x$ and ($x_1 \precsim x_n \precsim x$ for all n) and the combination of the two would complete the proof. We thus have $x_1 \sim y_1 \succsim x_n \sim y_n \succsim x \sim y$ for all n. By Lemma 2 we can select $\alpha_n \in [0,1]$ such that $\hat{y}_n \equiv \alpha_n y_1 + (1-\alpha_n) y \sim y_n$. By convexity of X^0, $\hat{y}_n \in X^0$. Thus, $u_d(x_n) = u_d(y_n) = u_d(\hat{y}_n)$ and $u_d(x) = u_d(y)$. Select a convergent subsequence $(n_k)_k$ such that $\hat{y}_{n_k} \to y^* \in X^0$. As u_d is continuous on X^0, we have $u_d(\hat{y}_{n_k}) \to u_d(y^*)$. Because $x_{n_k} \to x$ are in $X^{1\sim}$ and $\hat{y}_{n_k} \to y^*$ are in X^0, the two sequences are comparable and A2 implies that $x \sim y^*$ and thus also $y \sim y^*$. It follows that $u_d(x) = u_d(y) = u_d(y^*) = \lim u_d(\hat{y}_{n_k}) = \lim u_d(x_{n_k})$. $\square$

Let $v_* = \inf_{x \in X^{1\sim}} v(x)$ and $v^* = \sup_{x \in X^{1\sim}} v(x)$. Recall that v is bounded (by b, a) and we thus have $b \leq v_* \leq v^* \leq a$. The function u_d may or may not

be bounded (and it cannot be assumed bounded if it is supposed to represent P with a constant Δ). Denote $u^* = \sup_{x \in X^{1\sim}} u_d(x)$ and $u_* = \inf_{x \in X^{1\sim}} u_d(x)$ with the understanding that they can be ∞, $-\infty$, respectively.

On $X^{1\sim}$, both v and u_d represent $\succsim$, and both are continuous. This means that there exists a continuous, strictly increasing function $\psi : (v_*, v^*) \to (u_*, u^*)$ such that, for all $x \in X^{1\sim}$,

$$\begin{aligned} u_d(x) &= \psi(v(x)) \\ \lim_{v \searrow v_*} \psi(v) &= u_* \quad \lim_{v \nearrow v^*} \psi(v) = u^* \end{aligned}$$

Further, if v_* is obtained by v on $X^{1\sim}$, $u_* > -\infty$ and we can define $\psi(v_*) = u_*$, and, similarly, $\psi(v^*) = u^*$ in case $v^* = \max_{x \in X^{1\sim}} v(x)$ (and $u^* < \infty$).

We now wish to extend ψ to the entire range of v on X^1, preserving monotonicity and continuity. We first provide the obvious definition, and we then show that there exists $\gamma > 0$ such that, for every $x \in X^0$ and every sequence $(x_n) \subset X^1$ that converges to x, we have $u_d(x) - \lim u_d(x_n) = \gamma$.

Consider first $v > v^*$. If $X^{1\succ} = \varnothing$, then v on X^1 is bounded above by v^*, and the extension of ψ to this range is immaterial. Otherwise, that is, $X^{1\succ} \neq \varnothing$, we know, by A4, that $u_d(x) < \infty$ for all $x \in X^0$ and this means that $u^* < \infty$. We can set $\psi(v) = (v - v^*) + u^*$ for all $v > v^*$. This is a simple shift of the function v, which is obviously a continuous representation of $\succsim$ on $X^{1\succ}$. Similarly, consider $v < v_*$. If $X^{1\prec} = \varnothing$, then v on X^1 is bounded below by v_*, and the extension of ψ to this range is immaterial. Otherwise, that is, $X^{1\prec} \neq \varnothing$, we know, by A4, that $u_d(x) > -\infty$ for all $x \in X^0$ and this means that $u^* > -\infty$. Hence we can set $\psi(v) = (v - v_*) + u_*$ for all $v < v_*$.

Thus, $u_d(x) = \psi(v(x))$ is well defined for all $x \in X^1$; combined with the definition of u_d on X^0, we know that (i) u_d represents $\succsim$ on the entire space X; (ii) u_d is continuous on each of X^0 and X^1.

It remains to define $\gamma > 0$ and show that, for that γ,

$$u(x) = u_d(x) + \gamma \mathbf{1}_{\{x \in X^1\}} \tag{10}$$

is continuous on the entire space. We set γ to be equal to Δ as defined in Step 1.

Claim: $u : X \to \mathbb{R}$ defined in (10) is continuous on X.

Proof of the Claim: To verify continuity of u we only need to consider sequences $(x_n) \subset X^1$ that converge to $x \in X^0$ (because u_d is continuous on both X^0 and X^1, and X^0 is closed).

Let there be given such a sequence $(x_n) \subset X^1$ with $x_n \to x \in X^0$. Distinguish between two cases:

Case 2b(i): There exists $y \in X^0$ such that xPy.

In this case assume without loss of generality that $u(y) = u(x) - \gamma$, that is, that y is a u-maximal element with xPy. We know that there exists a sequence $(x'_n) \subset X^1$ with $x'_n \to x \in X^0$ and $x'_n \sim y$ so that, by the definition above, $u_d(x'_n) = u_d(y) = u(y)$ and, u_d on X^1 being a continuous transformation of v, where the latter is continuous on all of X, we also have $u_d(x_n) \to \lim_n u_d(x'_n) = u(y) = u(x) - \gamma = u_d(x) - \gamma$. Hence $u(x_n) = u_d(x_n) + \gamma \to u_d(x) = u(x)$ as required.

Case 2b(ii): There does not exist $y \in X^0$ such that xPy.

First observe that, by the definition of $u = u_d$ on X^0 in Step 1, we are in Case 1b and $u(x) = u_d(x) = v(x)$. Next, consider the sequence (x_n). Because it is convergent, and v is continuous on X, there exists $\lim_n v(x_n)$ (which is $v(x)$). On X^1 we have that $u_d(\cdot) = \psi(v(\cdot))$ is continuous, thus there exists $\lim_n u_d(x_n)$. The limit $v(x) = \lim_n v(x_n)$ cannot exceed v_* because if it did, there would be $y \in X^0$ such that $x_n \succsim y$ for infinitely many n's, and xPy would follow. However, in the domain $v \leq v_*$ we have $\psi(v) = (v - v_*) + u_* = v - (v_* - u_*)$. Further, in this case (corresponding to Case 1b in the definition of u on X^0), $u_* = \inf_{x \in X^0} v(x) = 0$ while $v_* = \inf_{x \in X^{1\sim}} v(x) = \Delta = \gamma$. It follows that

$$\lim_n u_d(x_n) = \lim_n \psi(v(x_n)) = \lim_n v(x_n) - \gamma = v(x) - \gamma$$

and thus $u(x_n) = u_d(x_n) + \gamma \to v(x) = u(x)$ and continuity is established.

□ ■

7.2 Proof of Corollary 1

We re-state axioms A6 in terms of the relation P and show that, when X is compact, it implies A4. First note that the axiom can be written as

A6 Lipschitz: There exists $\delta > 0$ such that, for every $x, y \in X^0$, if xPy then $\|x - y\| > \delta$.

Assume that X is compact, which implies that X^0 is compact as well. We wish to show that no infinite decreasing P chain can be bounded from below, nor can an infinite increasing P chain be bounded from above. However, A6 would make a stronger claim, namely, that there are no infinite P chains (neither increasing nor decreasing). Indeed, Lemma 12 implies that P is transitive. Had there been an infinite P chain, we would have to find two elements, say x_i and x_j such that $x_i P x_j$ (with $i > j$ for a decreasing P chain and $i < j$ for an increasing one) while they are in a δ-neighborhood of each other, in contradiction to A6. □

8 References

Arnot, C., P. C. Boxall, and S. B. Cash (2006), "Do Ethical Consumers Care about Price? A Revealed Preference Analysis of Fair Trade Coffee Purchases", *Canadian Journal of Agricultural Economics*, **54**: 555-565.

Arnould, E. J., and C. J. Thompson (2005), "Consumer Culture Theory (CCT): Twenty Years of Research", *Journal of Consumer Research*, **31**: 868-882.

Auger, P., P. Burke, T. M. Devinney, and J. J. Louviere (2003), "What Will Consumers Pay for Social Product Features?", *Journal of Business Ethics* **42**: 281-304.

Bajde, D. (2014), "Consumer Culture Theory: Ideology, Mythology and Meaning in Technology Consumption", *International Journal of Actor-Network Theory and Technological Innovation*, **6**: 10-25.

Barbier, E. B. (1993), "Sustainable Use of Wetlands Valuing Tropical Wetland Benefits: Economic Methodologies and Applications", *The Geographical Journal*, **159**: 22-32.

Barnes, J. I., C. Schier, and G. van Rooy (1997), "Tourists' Willingness to Pay for Wildlife Viewing and Wildlife Conservation in Namibia", DEA Research Discussion Paper no. 15.

Bartling, B., R. A. Weber, and L. Yao (2015), "Do Markets Erode Social Responsibility?", *The Quarterly Journal of Economics*, **130**: 219-266.

Basu, A. K. and R. L. Hicks (2008), "Label Performance and the Willingness to Pay for Fair Trade Coffee: A Cross-National Perspective", *International Journal of Consumer Studies*, **32**: 470-478.

Barnett, C., P. Cloke, N. Clarke, and A. Malpass (2005), "Consuming Ethics: Articulating the Subjects and Spaces of Ethical Consumption", *Antipode*, **37**: 23-45.

Bedate, A., L. C. Herrero, and J. A. Sanz (2004), "Economic Valuation of the Cultural Heritage: Application to Four Case Studies in Spain", *Journal of Cultural Heritage*, **5**: 101-111.

Beja, A. and I. Gilboa (1992), "Numerical Representations of Imperfectly Ordered Preferences (A Unified Geometric Exposition)", *Journal of Mathematical Psychology*, **36**: 426-449.

Ben-Porath, E., and I. Gilboa (1994), "Linear Measures, the Gini Index and the Income-Equality Tradeoff", *Journal of Economic Theory*, **64**: 443-467.

Bracha, A. and D. J. Brown (2012), "Affective Decision Making: A Theory of Optimism Bias", *Games and Economic Behavior*, **75**: 67-80.

Brunnermeier, M. K. and J. A. Parker (2005), "Optimal Expectations", *The American Economic Review*, **95**: 1092-1118.

Calabresi, G. (1985), Ideals, Beliefs, Attitudes, and the Law. Syracuse, NY: Syracuse University Press.

Calabresi, G. (2014), "Of Tastes and Values", Yale Law School Research Paper No. 500.

Chen, B. and X. Qi (2018), "Protest response and contingent valuation of an urban forest park in Fuzhou City, China", *Urban Forestry and Urban Greening*, **29**: 68-76.

De Pelsmacker, P., L. Driesen, and G. Rayp (2005), "Do Consumers Care about Ethics? Willingness to Pay for Fair-Trade Coffee", *Journal of Consumer Affairs*, **39**: 363-385.

Dillenberger, D. and P. Sadowski (2012), "Ashamed to be Selfish", *Theoretical Economics*, **7**: 99-124.

Duesenberry, J. S. (1949), *Income, Saving and the Theory of Consumer Behaviour.* Cambridge: Harvard University Press.

Eliaz, K. and R. Spiegler (2018), "A Model of Competing Narratives".

Enax L., V. Krapp, A. Piehl, and B. Weber (2015), "Effects of Social Sustainability Signaling on Neural Valuation Signals and Taste-Experience of Food Products", *Frontiers in Behavioral Neuroscience,* **9**: article 247 (1-15).

Evren, O. and S. Minardi (2015), "Warm-Glow Giving and Freedom to

be Selfish", *The Economic Journal*, **127**: 1381-1409.

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

Frank, R. H. (1985a), *Choosing the Right Pond: Human Behavior and the Quest for Status*. Oxford: Oxford University Press.

Frank, R. H. (1985b), "The Demand for Unobservable and Other Nonpositional Goods". *American Economic Review*, **75**: 101-116.

Garriga, E. and D. Mele (2004), "Corporate Social Responsibility: Mapping the Territory", *Journal of Business Ethics*, **53**: 51-71.

Glazer, J. and A. Rubinstein (2020), "Story Builders".

Hainmueller, J., M. J. Hiscox, and S. Sequeira (2015), "Consumer Demand for Fair Trade: Evidence from a Multistore Field Experiment", *The Review of Economics and Statistics*, **97**: 242-256.

He, H., Y. Li, and L. Harris (2012), "Social Identity Perspective on Brand Loyalty", *Journal of Business Research*, **65**: 648-657.

Heffetz, O. (2011), "A Test of Conspicuous Consumption: Visibility and Income Elasticities", *The Review of Economics and Statistics*, **93**: 1101-1117.

Hornsten, L. and P. Fredman (2000), "On the Distance to Recreational Forests in Sweden", *Landscape and Urban Planning*, **51**: 1-10.

Hotelling, H. (1947), "The Economics of Public Recreation", The Prewitt Report, National Parks Service, Washington.

Loureiro, M. and J. Lotade (2005), "Do Fair Trade and Eco-Labels in Coffee Wake Up the Consumer Conscience?", *Ecological Economics*, **53**: 129-138.

Kahneman D. and J. Knetch (1992), "Valuing Public Goods: The Purchase of Moral Satisfaction", *Journal of Environmental Economics and Management*, **22**: 57-70.

Karni, E. and Z. Safra (2002), "Individual Sense of Justice: A Utility Representation", *Econometrica,* **70**: 263-284.

Kreps, D. M. (1990), *A Course in Microeconomic Theory*. Princeton:

Princeton University Press.

Maccheroni, F., M. Marinacci, and A. Rustichini (2012), "Social Decision Theory: Choosing within and between Groups", *Review of Economic Studies*, **79**: 1591-1636.

Maccheroni, F., M. Marinacci, and A. Rustichini (2014), "Pride and Diversity in Social Economies", *American Economic Journal: Microeconomics*, **6**: 237-271.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995), *Microeconomic Theory*. Oxford: Oxford University Press.

Medin, D. L., H. C. Schwartz, S. V. Blok, and L. A. Birnbaum (1999), "The Semantic Side of Decision Making", *Psychonomic Bulletin and Review*, **6**: 562-569.

Morrison, M. (2002), "Understanding Local Community Preferences for Wetland Quality", *Ecological Management and Restoration*, **2**: 127-134.

Pigors M. and B. Rockenbach (2016), "Consumer Social Responsibility", *Management Science*, **62**: 3123-3137.

Prasad, M., H. Kimeldorf, R. Meyer, and I. Robinson (2004), "Consumers of the World Unite: A Market-Based Response to Sweatshops", *Labor Studies Journal*, **29**: 57-80.

Rubinstein, A. (1998), "Definable Preferences: An Example", *European Economic Review*, **42**: 553-560.

Ryan, R. M. (2001), "On Happiness and Human Potentials: A Review of Research on and Eudaimonic Well-Being", *Annual Review of Psychology*, **52**: 141-166.

Sheth, J., B. I. Newman, and B. L. Gross (1991), "Why We Buy What We Buy: A Theory of Consumption Values", *Journal of Business Research*, **22**:159-170.

Throsby, D. (2003), "Determining the Value of Cultural Goods: How Much (or How Little) Does Contingent Valuation Tell Us?", *Journal of Cultural Economics*, **27**: 275-285.

Varian, H. (1978), *Microeconomic Analysis*. London: Norton.

Veblen, T. (1899), *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions.* New York: Macmillan. Also: 1994 Dover paperback edition,1994 Penguin Classics edition.