

Tebbe, Eva; von Blanckenburg, Korbinian

Conference Paper

Does willingness to pay increase with the number and strictness of sustainability labels?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Environmental Economics: Consequences and Perceptions, No. A15-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Tebbe, Eva; von Blanckenburg, Korbinian (2017) : Does willingness to pay increase with the number and strictness of sustainability labels?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Environmental Economics: Consequences and Perceptions, No. A15-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168410>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Does willingness to pay increase with the number and strictness of sustainability labels?

Eva Tebbe^{*}, Korbinian von Blanckenburg[†]

Abstract

Labels signaling sustainable product attributes are gaining importance, although uncertainty concerning the environmental, micro- and macroeconomic benefits of such labels persist. One of the questions still incompletely answered is whether Willingness To Pay (WTP) varies with a gradually increasing number of labels on a food product. In order to answer this question, we conducted a laboratory experiment with 191 student respondents, testing consumer valuations of different labeling strategies. Using the Becker-DeGroot-Marschak mechanism, WTP for 15 food products was measured. The products were endowed with up to six different sustainability labels, such that each grocery item was available in eight product versions. For perishable, non-perishable and plant-based products, the results indicate that participants are prone to allocating WTP-premiums to labeled products, more than to unlabeled products. For animal products, however, labels do not influence WTP significantly. Furthermore, the premiums do not vary with an increasing number of labels, irrespective of whether the labels signal substitute or complementary sustainability information. The results are not entirely in line with normative notions of magnitude variation, but rather with the behavioral economic concept of embedding effects.

JEL Classification:

D11, D12, D44, L15

Keywords:

Behavioral economics; Experimental economics; Willingness to pay; Becker-DeGroot-Marschak mechanism; Organic; Embedding effect

Acknowledgements:

We thank Mirjam Lange, Benjamin Wegener and all participants of the 44th Hohenheimer Oberseminar as well as of the Vfs Annual Conference 2016 for their very helpful comments.

^{*} **Eva Tebbe, M.Sc.**

Research assistant, corresponding author
Ostwestfalen-Lippe University of Applied Sciences
Liebigstraße 87, 32657 Lemgo, Germany
eva.tebbe@hs-owl.de
00495261-7025122

[†] **Prof. Dr. Korbinian von Blanckenburg**

Professor for Economics and Mathematical Economics
Ostwestfalen-Lippe University of Applied Sciences
Liebigstraße 87, 32657 Lemgo, Germany

1 Introduction

Sustainability labels¹ are becoming increasingly common on supermarket shelves. They often do not occur individually, but a considerable number of products are endowed with two or even more of them. The rationale behind this is that consumers might be willing to pay more for products with more labels than for those without any or with fewer labels, assigning value to each label individually and accumulating the associated information. However, the functionality of labels as instruments signaling valuable information may be limited. Some research indicates that consumers neither thoroughly trust labels nor fully understand them (Harper and Makatouni, 2002; Pedersen and Neergaard, 2006; Sirieix et al., 2013). Closely connected to this, it is also questionable whether premiums for individual labels accumulate if several labels are presented together on a product. Studies dealing with this question regularly identify substitution and complementary effects between different labels, depending on consumer associations with these particular labels, as well as on other product attributes (Dufeu et al., 2014; Gracia et al., 2014; Onozaka and Mc Fadden, 2011). Yet, there are also findings supporting behavioral economic concepts like embedding or information overload effects and attitude-behavior gaps (Bougherara et al., 2007; Irwin and Spira, 1997; Jongmans et al., 2014).

The question to be answered in the current research is as follows: Are consumers willing to pay a premium for food products endowed with sustainability labels, and if so, does it vary with an increasing number of labels? Specifically, this paper considers how WTP develops if up to six sustainability labels are gradually added on a variety of products. Among the tested labels are five organic ones with a differing stringency of certification criteria. Additionally, a fair trade label is included in the analysis, so as to draw more precise conclusions concerning the adequacy of normative versus behavioral economic concepts. As the market for sustainable food products is among the most developed sustainable markets, and as food purchase frequency is generally high, this area is well suited to learning about consumer sensitivity to and perceptions of sustainability labels. Previous research in this area focusses mainly on individual or a small number of products. However, there might be patterns of valuation among consumers with regard to sustainable food products in general, or regarding broader product groups, which have not been elaborated so far. Extending the number of analyzed groceries used in other studies, ours considers 15 different products, the WTP for which is analyzed in an all-encompassing manner, as well as differentiated with regard to perishable, non-perishable, plant-based and animal products. Accordingly, our paper differs methodologically from previous studies not only in terms of the number of labels tested, but also with regard to the number of products. We introduce a relative measure of WTP called Aggregated Standardized Relative WTP (ASRWTP) which enables comparing different products.

Concerning the experimental valuation methodology, we follow benchmark practices developed over the last 20 years (Alfnes and Rickertsen, 2011; Lusk and Shogren, 2007). In order to control for as many factors as possible, the study is conducted as a laboratory experiment (Alfnes and Rickertsen, 2011; Guala, 2000). Furthermore, the Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964) is chosen instead of hypothetical statements for revealing WTP, thus eradicating hypothetical bias. Besides motivating our participants by providing real economic incentives directly connected to their bid-

¹The European Union (2015) defines sustainable development as a means to meet “[...] the needs of present generations without jeopardizing the ability of futures generations to meet their own needs – in other words, a better quality of life for everyone, now and for generations to come. It [...] regards social, economic and environmental issues as inseparable and interdependent components of human progress.” Labels signaling certain aspects fostering sustainable development can be called sustainability labels. Thus, organic and fair trade labels can be classified as sustainability labels, although sustainability might not be the only aspect associated with these labels by consumers (indeed, e.g., they might also perceive organic labels as healthier).

ding behavior as it is standard with the BDM mechanism, we also introduce a conundrum in our experimental design. This is a question which obviously does not make sense and its detection is connected to additional monetary outcomes in the experiment. This design serves as an incentive for our participants to stay concentrated throughout the experiment.

2 Related literature

Sustainability labels are intended to serve as signaling instruments for sustainable product attributes. A significant amount of research analyses the effects of such labels on consumer product valuation. The basic findings can be summarized as follows. First, adding a label to a previously unlabeled product – that is, adding signals of one or more sustainability attribute(s) – usually enhances customer WTP (Costanigro et al., 2014; Loureiro et al., 2001). Second, there are heterogeneous consumer segments when it comes to the valuation of sustainability labels (Kehlbacher et al., 2012; Loureiro et al., 2001; Loureiro et al., 2002; Loureiro and Lotade, 2005; Pelsmacker et al., 2005). Third, overall product characteristics, like brand and taste, seem to influence the effects of individual sustainability labels (e.g. Bauer et al., 2013; Sörqvist et al., 2013). Fourth, as labels are only imperfect sustainability signals, trust and knowledge appear to be relevant for the evaluation (Costanigro et al., 2014; Sirieix et al., 2013). Above and beyond this, questions concerning the influence of the number and combination of individual labels arise, and the number of studies dealing with the effects of multilabeling is growing (e.g., Costanigro et al., 2014; Dufeu et al., 2014; Jongmans et al., 2014; Sirieix and Tagbata, 2008; Sirieix et al., 2013). Theoretically, following a notion of magnitude variation raised by normative theory, if single sustainability labels yield a percentage increase on consumer WTP, more labels with additional sustainability information should yield an even greater effect. On the other hand, more labels signaling the same information would surely not increase consumer WTP (Fischhoff et al., 1993).

However, behavioral decision theory suggests that there may be counter-intuitive effects violating the rule of normative theory, one of which is called the “embedding effect”. Until now, this phenomenon is found predominantly in contingent valuation studies. It characterizes situations in which the joint presentation of several labels signaling different information yields the same premium as a minor subset of the labels, because consumers only obtain utility from “sustainability” as such, but no additional value if there is “a higher degree of sustainability” (Einhorn and Hogarth, 1981; Fischhoff et al. 1993; Irwin and Spira, 1997; Jongmans et al., 2014; Kahneman and Knetsch, 1992).

In a like manner, the phenomenon of information overload induces similar results (i.e., premiums do not rise with an increasing number of labels). If the cost of processing additional information provided by another label is higher than the expected utility from internalizing it, consumers might disregard that extra information (Bougherara et al., 2007; Dufeu et al., 2014; Eppler and Mengis, 2008).

As a third anomaly of consumer behavior, divergences between attitudes and actual behavior, driven by a lack of environmental awareness, emotional involvement or locus of control, are regularly discovered (Irwin and Spira, 1997; Kollmuss and Agyeman, 2002; Luchs et al., 2010; Vermeir and Verbeke, 2006). Bearing these phenomena in mind, interpreting empirical findings becomes precarious.

Diving into raw research results, studies yield varying findings on the existence, direction and extent of the effect of multilabeling patterns on WTP. Sirieix and Tagbata (2008) tested WTP for organic and fair trade chocolate of 102 subjects in an experimental framework, using the BDM mechanism. They deduce that, compared to single labels, combining social and environmental labels on one product leads to a sub-additivity of WTP. Costanigro et al. (2014), however, find, in their experimental auction of local and organic apples with 109 participants, that multiple labels induce a disproportionate increase in WTP if the visible information is combined with good taste (literally). Dufeu et al. (2014) compare contingent valuation with the BDM mechanism for eliciting WTP for honey with up to three quality, organic and

fair trade labels of 519 customers in online surveys and face-to-face-interviews. They show that generally, increasing the number of labels increases WTP. This relationship is relatively linear with contingent valuations, whereas the marginal increase in value decreases in the case of the BDM procedure. The current paper contributes to the abovementioned research on the interaction between an increasing number of sustainability labels as signals of sustainable product attributes and WTP and adds to the current state of the art in two respects. First, this study analyzes a broader variety of products than previous research. In addition to an overall analysis including all products, it also deploys nuanced analyses considering four groups: plant-based, animal, perishable and non-perishable products. This provides a departure from specific products, so as to elaborate whether there might be product-independent or product group specific valuation-patterns. Second, the number of labels is enhanced compared to other work, while the labels vary with regard to certification criteria, yielding a more precise map of labeling effects on WTP for sustainable products. The label-composition considered for the study at hand consists of five organic labels with different levels of certification strictness and one fair trade label.

3 Experimental design and procedure

3.1 Methodological setting

In terms of the methodological framework, this study is conducted as a laboratory experiment. Although laboratory experiments are criticized for their artificial environment, in the context of sustainability labels, at the current point of research, the method is appropriate, due to scientific results indicating many factors influencing WTP for sustainable products, some of which have not yet been fully identified and allocated (Kollmuss and Agyeman, 2002; van Huylenbroek et al., 2009). Consequently, controlling for as many influences as possible is important.

Regarding the evaluation of bids for food products, in the context of sustainability, there is strong evidence of discrepancies between stated and actual behavior (Chang et al., 2009; Cummings et al., 1995; Irwin and Spira, 1997). Thus, mechanisms using real economic incentives to measure homegrown WTP values should be preferred over methods eliciting hypothetical WTP (Alfnes and Rickertsen, 2011). For eliciting WTP for sustainable products in the laboratory, various mechanisms are used in the literature. Common ones include forms of sealed bid n th-price auctions, such as the Vickrey auction for $n = 2$ (Vickrey, 1961; e.g., used by Bernard et al., 2006; Bernard and Bernard, 2009; Bernard and He, 2010; Bernard and He, 2011; Costanigro et al., 2014), as well as English auctions (e.g., Lusk et al., 2004) and the BDM mechanism (Becker et al., 1963; e.g., used by Dufeu et al., 2014; Sirieix and Tagbata, 2008; Wertenbroch and Skiera, 2002). The latter has been shown to be well-suited for eliciting consumers' real WTP and to be close to actual shopping behavior (Alfnes and Rickertsen, 2011; Becker et al., 1964; Bohm et al., 1997; Guala, 2000). Moreover, it is independent of the number of participants (Alfnes and Rickertsen, 2011). Accordingly, and in line with related research conducted by Dufeu et al. (2014), Shi et al. (2013) and Sirieix and Tagbata (2008), the study at hand is based on a BDM laboratory experiment.

3.2 Product and label selection and treatment setup

To resemble real world purchasing decisions as well as customary product packaging as closely as possible, it was necessary to select (a) product types as well as, in the case of packaged food, (b) existing product brands prior to designing the final experimental layout.

- (a) The selection process of the tested products balanced perishable and non-perishable as well as animal and plant-based products. Products were chosen on the basis of an analysis of the Harmonized Index of Consumer Prices (HICP), in such a way that nine out of ten food product categories of the

HICP are represented by (at least) one of the products tested.² As a result, there were 15 products chosen (in customary sizes) (Table 1).

- (b) To resemble real world conditions closely, for the packaged goods, existing brands were chosen. For the purpose of selecting brands, a preliminary online survey was conducted. This survey asked 166 respondents about their price assessment, their knowledge as well as their impression of and attitude towards several product brands for each of the selected packaged products. The repertoire included both high- as well as medium- and low-end brands. Based on the results of this survey, for each of the packaged goods, a brand was chosen which most participants did not know and which was evaluated moderately in terms of appeal and price-level. These criteria were implemented to reduce potential bias due to knowledge, brand or market price effects, as previous studies show that these factors influence consumer decisions (Pelsmacker et al., 2005; Degeratu et al., 2000; Devinney et al., 2010; Muller and Ruffieux, 2011).

Table 1: Categories of selected food products and sizes.

	Animal products	Plant-based products
Perishable	Fresh Rump steak (250 g) Fresh Chicken breast (300 g) Smoked salmon (100 g)	Fresh Apples (1 kg) Fresh Bananas (300 g) Fresh Lemons (2 pieces) Fresh Basil (1 pot) Fresh Tomatoes (500 g)
Non-perishable	Long-life milk (1 liter) Canned tuna (120 g) Milk chocolate (100 g) Bee honey (500 g)	Ground coffee (500 g) Rolled oats (500 g) Packaged sunflower bread (500 g)

Besides the product- and brand-selection, another important step for preparing the experiment was to decide on the labels to be placed on the products. Concerning ecological issues, there are five third-party labels characterizing the German market for sustainable products: The European Union's organic logo ("Euro-leaf", EU), the German organic label ("Bio-Siegel", GER), the Bioland (BL) as well as the Naturland (NL) labels for organic agricultural production and Demeter (DM), a label for bio-dynamic agriculture. Additionally, as it is particularly suitable for testing complementary effects of different labels on WTP, a popular fair trade label (FT) was added to the collection of labels to be analyzed. In contrast to the other labels, the fair trade label predominantly represents altruistic aims, as the customer her-/himself does not benefit directly from higher salaries for producers. Organic labels, however, are often associated with private benefits, such as higher quality in general, higher nutritional value and less pesticides. Table 2 categorizes the depicted labels.

For testing the effects of an increasing number of labels, the experiment deployed WTP for each product with different label combinations, as well as for an unlabeled product version of each grocery item. Because with six selected labels, the number of all possible label combinations is large³, it was not practical to test all potential combinations. As a consequence, we restrict our design to specific label combinations which allow for an analysis of the effects of a successively increasing information load on WTP. In order to depict these effects, the number of labels increases with each successive label combination. Besides the unlabeled product version, seven further product versions, differentiated by seven label combinations, were created. Accordingly, there are eight product versions v (with $v =$







² European Central Bank. Measuring inflation – the Harmonised Index of Consumer Prices (HICP). Available at: https://www.ecb.europa.eu/stats/prices/hicp/html/hicp_coicop_inx_2016-05.en.html. The only category we do not consider in this study is cooking fat and oils, which accounts for only 2,6% of the total amount of products given in the HICP.

³ Six labels yield 63 possible label combinations (excluding the unlabeled product version). With 15 products, this would result in 945 testable product-label combinations, which is beyond the scope of this study. Accordingly, we restrict our analysis to a subsample of label combinations.

0; 1; ...; 7) of each product, with $v = 0$ representing the unlabeled product version. The order of added labels was based on three aspects:

- 1) *Obligation*: The EU label is obligatory for pre-packaged organic products in the European Union and thus has to be placed on all products following the European Union's standards for organic agriculture. Consequently, none of the other labels can be displayed without displaying the EU label. The EU label, hence, represents the first label combination ($v = 1$).
- 2) *Public and private labeling schemes and strictness of certification criteria*: The EU logo as well as the GER label refer to the European Union's statutory provisions of organic agriculture, thus, representing exactly the same aspects, whereas BL, NL and DM are awarded by private organizations defining their own standards. BL as well as NL labels require more stringent conditions than the EU as well as the GER label, and terms for the DM logo are even stricter. Accordingly, the combination of the GER and the EU label represents the second label combination ($v = 2$). For the third label to be added, both BL and NL qualify equally, as they define comparably strict (yet slightly different) criteria. Thus, two label combinations with three labels were designed ($v = 3$ and $v = 4$), as well as a combination including both ($v = 5$). With both NL and BL labels, DM is added in the sixth label combination ($v = 6$), as it is the only organic label classified as "bio-dynamic", which represents the most stringent production standards.
- 3) *Representation of sustainability aspects*: As all the abovementioned labels focus primarily on ecological agriculture and are partly associated with private aims, the FT label expresses another dimension of sustainability, namely a social and altruistic one. This label is added in the seventh label combination for each product type ($v = 7$).

Table 2: Basic characteristics of six sustainability food labels.

Label (abbr.)	Sustainability food label	Category	Obligation	Public vs. private	Certification organization
EU	EU "Euro-leaf" 	Organic	Obligatory for pre-packaged organic food	Public	European Union
GER	German "Bio-Siegel" 	Organic	Voluntary	Public	German Federal Ministry of Food and Agriculture
BL	Bioland label 	Organic	Voluntary	Private	Bioland Association (Bioland-Verband für organisch-biologischen Landbau e.V.)
NL	Naturland label 	Organic	Voluntary	Private	Naturland Association (Naturland-Verband für ökologischen Landbau e.V.)
DM	Demeter label 	Organic and biodynamic	Voluntary	Private	Demeter Association (Demeter e.V.)
FT	Fair trade label 	Fair trade	Voluntary	Private	Fair Trade Association (TransFair Verein zur Förderung des Handels mit der "Dritten Welt" e.V.)

The eight product versions v (with $v = 0; 1; \dots; 7$) are presented in Table 3.⁴ Analogous to the approaches of Dufeu et al. (2014), Jongmans et al. (2014) Larceneux et al. (2012), Pelsmacker et al. (2005)

⁴ For the respective certification criteria, see http://ec.europa.eu/agriculture/organic/index_de.htm (EU); http://www.bmel.de/DM/Landwirtschaft/Nachhaltige-Landnutzung/Oekolandbau/_Texte/Bio-Siegel.html (GER);

and Sénéchal et al. (2014), pictures of the selected groceries were presented during the experiment. To be able to generate eight product versions for each grocery, we needed to modify the product-pictures. We only sold those product versions that actually existed.⁵ Apparently, with this design, not all labels are treated independently, so that it is not possible to make statements about the effect of individual labels, independent of the other labels. However, the design chosen here seems appropriate for testing a labeling effect as well as a multilabeling effect, focusing on the overall information-load a certain label combination delivers.

Table 3: Label combinations of the tested product versions.

Product version	Version code (v)	Label					
		EU	GER	NL	BL	DM	FT
Unlabeled	0	-	-	-	-	-	-
One label	1	x	-	-	-	-	-
Two labels	2	x	x	-	-	-	-
Three labels (NL)	3	x	x	x	-	-	-
Three labels (BL)	4	x	x	-	x	-	-
Four labels	5	x	x	x	x	-	-
Five labels	6	x	x	x	x	x	-
Six labels	7	x	x	x	x	x	x

Table 4 specifies, for each comparison between two product versions, whether switching from the lower version number to the higher product version should yield an increase in utility on the consumer side, following normative concepts. This holds if and only if the labels included in the higher version number signal complementary information compared to the labels of the lower version number. For example, the marginal utility from switching from version 1 to version 2 should be zero, as the two versions signal exactly the same information. However, all other versions differ with regard to the values presented, so that complementary effects can be expected.

Table 4: Substitution and complementary effects of the product versions.

Product version ^a	Unlabeled ($v = 0$)	One label ($v = 1$)	Two labels ($v = 2$)	Three labels (NL) ($v = 3$)	Three labels (BL) ($v = 4$)	Four labels ($v = 5$)	Five labels ($v = 6$)
One label ($v = 1$)	Complementary	Substitution	Complementary	Complementary	Complementary	Complementary	Complementary
Two labels ($v = 2$)	Complementary						
Three labels (NL) ($v = 3$)	Complementary						
Three labels (BL) ($v = 4$)	Complementary						
Four labels ($v = 5$)	Complementary						
Five labels ($v = 6$)	Complementary						
Six labels ($v = 7$)	Complementary	Complementary	Complementary	Complementary	Complementary	Complementary	Complementary

^aVersion codes, as defined in Table 3, are given in parenthesis.

The table indicates whether the product version given in the row includes labels signaling substitutable or complementary values, compared to the product version given in the column.

<http://www.bioland.de/start.html> (BL); http://www.naturland.de/startseite_naturland.html (NL); <http://www.demeter.de/> (DM); <http://www.fair-trade-deutschland.de/index.php> (FT).

⁵ Colson et al. (2016) show that researchers and students consider selling mislabeled products objectionable, but that displaying manipulated product-pictures is regarded less objectionable.

With 15 products and eight tested label combinations, there were 120 product-label combinations. In order to test them all, a crossover experimental design was chosen, so that the product-label combinations were allocated to eight treatment groups. Each product type was presented in only one version within each treatment group, except coffee, which occurred at the very beginning and the very end of the sequences in two different versions, so as to allow for manipulation checks. Consequently, participants had to conduct 16 evaluations. Among the treatment groups, the product versions differed, while the product-sequence was kept the same, so that groups are differentiated according to label combinations only. The product versions were allocated among the eight treatment groups in such a manner that in each group, for each of the eight label combinations, two products were presented. Table 5 depicts the specific product versions which were presented in the eight treatment groups.

Table 5: Product versions within the eight treatment groups.^b

Product	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Ground coffee	0	1	2	4	3	5	6	7
Lemons	6	7	0	1	2	4	3	5
Sunflower bread	1	2	3	0	5	6	7	4
Apples	4	3	5	6	7	0	1	2
Chicken breast	1	2	4	3	5	6	7	0
Milk chocolate	5	6	7	0	1	2	3	1
Long-life milk	6	7	0	1	2	4	3	5
Bee honey	4	3	5	6	7	0	1	2
Rolled oats	3	5	6	7	0	1	2	4
Rump steak	0	1	2	4	3	5	6	7
Smoked salmon	7	0	1	2	4	3	5	6
Tomatoes	2	4	3	5	6	7	0	1
Canned tuna	2	4	3	5	6	7	0	1
Basil	5	6	7	0	1	3	4	2
Bananas	3	5	6	7	0	1	2	4
Ground coffee	7	0	1	2	4	3	5	6

^b The numbers refer to the specific product version as defined by version codes in Table 3. A value of 7 in Table 5 thus refers to $v = 7$, which is the version with six labels. The order of the products displayed in the “products”-column represents the order of the products in all treatment groups. For example, treatment group 1 started with the presentation of ground coffee in the 0-label-version, continued with lemons in the 6-label version, and sunflower bread in the 1-label-version was presented third.

3.3 Procedure

The experiment took place at the Ostwestfalen-Lippe University of Applied Sciences in Lemgo, Germany, in spring 2015. 191 participants, mainly students, were recruited through online advertisements and invitations. Participants neither knew what the experiment was about nor had taken part in comparable tasks before. The laboratory procedure followed seven steps:

First, after the participants took seats in front of separated and well divided computers, they were asked to start a short explanatory video, which described the experimental procedure and informed about the financial compensation for participation. This video described the procedure neutrally, so as to reduce social experimenter demand effects (see Zizzo (2010) for details about these effects). The video pointed out that answers would be kept anonymous to obviate bias induced by social desirability (Brown and Cohen, 2015) and that one of the presented products would be auctioned at the end of the experiment. The decision to sell only one product was important so as to ensure that participants’ WTP did not decline towards zero with an increasing number of products shown, as they might be wary of buying too much or spending too much money. Participants did not know which product was selected until the experiment was over. Moreover, the video explained that there would be a conundrum included in the series of experimental questions and that each participant who identified the conundrum would obtain an additional random monetary premium. The conundrum is a question which did obviously not make sense and not fit into the series of questions asked. Attentive participants should easily spot it. To the

best of our knowledge, we are the first to choose a conundrum to control for whether participants concentrate throughout the experiment. In a second step, the subjects were prompted to answer a short self-test to check whether they understood the BDM mechanism explained in the video. Third, a demographic and lifestyle questionnaire was to be answered and fourth, evaluation rounds started. The experiment was implemented in z-Tree (Fischbacher, 2007). For each of the 16 product-label combinations, subjects had to type in their maximum WTP so that there were 16 experimental rounds per treatment. After typing in their WTP, subjects had to answer a series of questions concerning their attitudes towards the specific product and about their consumption habits in terms of the product type. Furthermore, they were asked for the average price they usually pay for a conventional version of the product in the marketplace. This value serves as an individual-specific reference market price for a conventional product version for calculating the relative WTP-premium as defined in the following Section 4. After the evaluation rounds were over, two further questionnaires had to be answered, which asked about shopping behavior, attitudes concerning sustainable products, knowledge and assessment of the specific labels. As a sixth step, one of the 15 products was announced as to be sold and a random price determined. In each group, a product was deliberately chosen which exists as presented during the experiment. To avoid participants anticipating the product for sale, each group received a different one. The price range was from one third of the actual market price to triple the market price. Following Bohm et al. (1997), the participants were not informed about this price range to prevent upper bound adaptations of stated WTP. Finally, participants were paid individually. Those who bid at least as high as the random price received the product and were equipped with the starting capital (9 euros⁶) minus the random product-price, plus a contingent premium (ranging from 0.50 to 2.50 euros) for identifying the conundrum correctly. The others received 9 euros plus the extra premium if they tagged the conundrum right.

4 Conceptualization of the dependent variable

As derived in Section 3.2, our paper differs from previous research particularly with respect to the number of labels and the number of products included in the study, allowing for an analysis which is independent of individual products. In order to conduct such an analysis, instead of absolute numbers, the research at hand measures WTP in terms of the percentage premium an individual is willing to pay for a certain food product version (as defined in Table 3), compared to the price he or she usually pays for a conventional version. This calculation corrects WTP by the subjectively estimated price for a standard product version and thereby controls for product differences and anchoring effects. Moreover, it controls for unobserved general heterogeneity in the valuation of specific food products within the sample. The procedure is in line with the findings of Muller and Ruffieux (2011) and Bernard and He (2010), who show that WTP is biased by estimated and real market prices, as well as with the results from Shi et al. (2013), who provide evidence of purchase intention influencing full WTP bids. Before being able to analyze the data, the stated absolute WTP thus has to be transformed into a relative measure, which we call Standardized Relative WTP (SRWTP). Equation (1) estimates the SRWTP which an individual is willing to pay for a certain product version of a specific grocery.

$$SRWTP = (bid - stated\ price) / stated\ price \quad (1)$$

The bid is interpreted as the absolute WTP of a subject for a specific product version of a certain grocery item. The stated price is the market price for a conventional product version as estimated by the same individual, i.e. the price he or she usually pays for this kind of grocery.

As with SRWTP, we now have a relative measure and are able to pool products in one variable in a second step. We call this pooled variable Aggregated SRWTP (ASRWTP), which depicts the distribution of SRWTP for a specific product version, aggregated among all product types. Consequently,

⁶ 1 euro = 1.08 USD in April 2015.

ASRWTP is a product-independent variable, displaying standardized percentage WTP-premiums or discounts for a certain product version, compared to the prices subjects usually pay in the marketplace. As far as we know we are the first to move away from individual products and to introduce ASRWTP as an independent measure.

5 Hypotheses

Paralleling descriptions in Section 2, studies predominantly support the view that labels increase WTP, compared to unlabeled products (Loureiro et al., 2001; Loureiro et al., 2002; Sirieix and Tagbata, 2008; Sörqvist et al., 2013), raising the question of whether these results prove valid for the sample treated in this study. Accordingly, the first null hypothesis deals with the existence of a labeling effect.

H0₁: Labeling Effect: *ASRWTP is not higher for labeled food products compared to unlabeled food products.*

Concerning products with more than one label, there are controversial theoretical concepts and empirical findings. In general, it appears decisive to determine whether normative concepts of magnitude variation represent real consumer decisions. According to this notion, if certain labels represent identical product attributes, the marginal effect of adding all these labels, compared to only one of them or a certain (strict) subset, should be zero (Dufeu et al., 2014). The second null hypothesis is therefore:

H0₂: Substitution Effect: *ASRWTP is not higher for food products labeled with more labels, compared to food products with fewer labels, if the added labels do not represent additional product values.*

However, for labels representing different sustainability attributes, i.e., certification criteria of varying strictness, standard random-utility models indicate that there may be complementary effects, leading to increasing WTP the more labels that are added (Dufeu et al., 2014). Thus, the third null hypothesis is:

H0₃: Complementary Effect: *ASRWTP is not higher for food products labeled with more labels, compared to food products with fewer labels, if the added labels do represent additional product values.*

The three hypotheses focus on the effect of labels. However, our design includes various products from different product groups and there might be inter-group differences regarding WTP-premiums. Yet, it is an open question how WTP depends on the product group to which a certain grocery item belongs. Some evidence with regard to this question is delivered by Bernard and He (2011), who analyze two pairs of fresh and processed foods and show that the latter have a higher substitutability among conventional, non-GM and organic product versions than fresh ones. However, fresh organic or non-GM products did not necessarily yield higher WTP than processed versions. To the best of our knowledge, there is no study which differentiates between non-perishable and perishable products, as well as between animal and plant-based products. Thus, as we do not have a testable hypothesis concerning inter-group differences, we formulate an Exploratory Research Question as follows:

Exploratory Research Question: Differences between four product groups

Is there a difference in ASRWTP for one or more labels between non-perishable, perishable, animal and plant-based products?

6 Results

Concerning the sample characteristics of the 191 subjects taking part in the experiment, roughly 39% are female and 61% are male. The average age is 24 and most of the participants are neither married nor have children (93% each). More than three quarters of the sample completed secondary school, but only 7% have a Bachelor's degree. The sample is relatively homogenous concerning demographics, which is due to methodological and organizational factors, as laboratory experiments at universities regularly yield such groups. 87% of the participants shop for food at least once a month and 68% cook more than

4 times per month. Moreover, nearly all participants are neither vegan nor vegetarian (97%) and most frequently shop at supermarkets or discounters (ca. 90%), while the remaining subjects predominantly purchase food at local farmers' markets, wholefood or health shops. 62% of the respondents consider organic labels as important or even as very important. Regarding fair trade labels, 73% fall in this category. The percentage of participants who at times, regularly or exclusively buy organic is 43%, while the respective share is 38% for fair trade food products.⁷ This pattern corresponds to the spread of subjects among categories of label-specific variables. More than half the respondents neither know well nor recognize the NL and DM labels, while most are familiar with the GER and the FT labels and notice them often when shopping. As organic and fair trade products available at supermarkets and discounters usually have these two well-known labels, but rarely conform to NL or DM standards, these distributions match the shopping habits of the sample as a whole. The EU as well as the BL label are vaguely known by most participants and noticed regularly, but not as often as the GER and the FT label. For the EU logo, this is surprising, as it is both common and obligatory for all pre-packaged organic products. A possible explanation is that the recognition of the EU label might be dominated by the GER label, which is regularly added to the EU label since it does not require fulfillment of additional regulatory criteria and is often placed at more popular positions on product packages. Due to its design containing the word "bio", the GER label can easier be associated with organic standards. In terms of label values, it is noteworthy that a considerable proportion of the sample associates all labels except the DM (which most participants do not know) and the FT label not only with environmental friendliness and being free of chemicals, but also with higher product quality, healthier food and higher food safety than conventional products. Accordingly, they might add WTP-premiums for these labels not only because of sustainability-concerns, but also because of expected health or taste benefits. For the FT label, associated values are mainly social, that is, supporting local farmers and fair working conditions.

Table 6: Descriptive statistics of the dependent variables ASRWTP^c.

Product version ^d	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Obs. ^e
Unlabeled ($v = 0$)	-1.0000	-0.2000	0.0000	0.0454	0.2500	5.0000	241
One label ($v = 1$)	-1.0000	-0.0935	0.0256	0.1424	0.3333	4.4000	263
Two labels ($v = 2$)	-1.0000	-0.1233	0.0000	0.1877	0.4133	5.5220	268
Three labels (NL) ($v = 3$)	-1.0000	-0.1133	0.0871	0.1575	0.3469	4.0000	264
Three labels (BL) ($v = 4$)	-1.0000	-0.1429	0.0000	0.1716	0.3333	4.0000	245
Four labels ($v = 5$)	-1.0000	-0.1000	0.0020	0.2127	0.3550	14.000	251
Five labels ($v = 6$)	-1.0000	-0.0980	0.1150	0.1672	0.5000	3.5000	254
Six labels ($v = 7$)	-1.0000	-0.0104	0.1445	0.2288	0.4170	5.0000	260

^c ASRWTP depicts the Aggregated Standardized WTP, that is the marginal WTP a person is willing to pay for a certain label combination, compared to conventional products in the supermarket.

^d Version codes, as defined in Table 3, are given in parenthesis.

^e With 191 subjects tested and 16 WTP-statements of each participant, 2046 observations of absolute WTP for different product-label combinations were generated, after clearing up observations corresponding to participants who are constantly nonresponsive (as advised by Alfnes and Rickertsen (2011)) and WTP corresponding to subjects who stated that they assume the market price for a conventional product version to be zero.

To investigate how the above described attitudes turn into WTP-premiums, we analyze our dependent variable ASRWTP. The descriptive results in Table 6 show that there are outliers in the sample, which induces us to focus our analysis on the more robust median rather than mean bids. Median ASRWTP varies across product versions, while it is zero for unlabeled products, as well as for product version 2

⁷ In 2014, the share of German people who exclusively, predominantly or partly buy organic was 35.47% (<http://de.statista.com/statistik/daten/studie/172357/umfrage/einkaufsmenge-bioproducte/>). Thus, compared to the overall German population, the share of organic customers is roughly 7.5% higher in the present sample. However, Clamor (2010) shows that younger people with a higher educational level are more prone to buy organic products regularly.

and product version 4. Median ASRWTP is highest for product version 7, indicating a willingness to pay a premium of 14.45% for products endowed with six labels.

Above and beyond this first glance at the descriptive statistics, the two-sided non-parametric Brunner-Munzel Test (Brunner and Munzel, 2000) was performed to test the statistical significance of differences between ASRWTP for different label combinations. Using these test statistics, the distribution of ASRWTP for each label combination is compared to the ASRWTP of each other label combination. Results are shown in Table 7. The generated p-values testify a significant difference between unlabeled and labeled products. However, there are no significant differences between any labeled products. Hence, H_{01} can be rejected, as all product versions with at least one label increase ASRWTP significantly, compared to unlabeled versions. Furthermore, the results suggest that for all label combinations, WTP is inelastic to the number of labels, independent of whether they provide complementary or substitute information. Thus, we cannot reject H_{02} and neither H_{03} .

Table 7: Test statistics of the two-sided Brunner-Munzel Test.^f

Product version ^g	Unlabeled ($v = 0$)	One label ($v = 1$)	Two labels ($v = 2$)	Three labels (NL) ($v = 3$)	Three labels (BL) ($v = 4$)	Four labels ($v = 5$)	Five labels ($v = 6$)
One label ($v = 1$)	2.4579** (0.0143)						
Two labels ($v = 2$)	2.3723** (0.0180)	0.1163 (0.9075)					
Three labels (NL) ($v = 3$)	2.6263*** (0.0089)	0.2706 (0.7868)	0.0584 (0.9535)				
Three labels (BL) ($v = 4$)	2.1646** (0.0309)	-0.1925 (0.8475)	-0.2266 (0.8208)	-0.3998 (0.6895)			
Four labels ($v = 5$)	2.1896** (0.0290)	-0.1095 (0.9128)	-0.1427 (0.8866)	-0.3426 (0.7320)	0.0242 (0.9807)		
Five labels ($v = 6$)	3.2570*** (0.0012)	1.0676 (0.2862)	0.8646 (0.3877)	0.8006 (0.4237)	1.1139 (0.2659)	1.1026 (0.2707)	
Six labels ($v = 7$)	3.8189*** (0.0002)	1.4782 (0.1400)	1.2205 (0.2228)	1.1585 (0.2472)	1.5482 (0.1222)	1.4955 (0.1354)	0.2134 (0.8311)

^f p-values are reported in parenthesis.

^g Version codes, as defined in Table 3, are given in parenthesis.

*** Significant at the 1% level; ** Significant at the 5% level.

The rationale behind this is as follows. Concerning H_{02} , full substitution could be expected for the EU and the GER label. In fact, adding the GER label does not yield significantly higher bids, compared to the EU label only. Thus, it might be that these labels are considered substitutes. But even though the notion of the existence of substitution effects cannot be rejected, it is not capable of explaining all missing differences between the product versions. As soon as the sustainability attributes stated on the products do not solely contain naturally occurring correlations, consumers should value them separately, if their valuation follows normative principles (Irwin and Spira, 1997). Accordingly and in contrast to H_{03} , normative concepts would suggest adding a label which signals complementary information compared to the present labels would justify a WTP-premium if subjects know about the difference between the labels. In our case, most subjects state that the FT logo signals different values compared to the organic labels, but it does not motivate them to pay a significant premium. Assuming that this might be due to the altruistic nature of the fair trade values, comparisons among different certification-standards of the organic labels could provide further support for the non-existence of complementary effects. Comparing organic and bio-dynamic products, there is no difference between the 5- or the 6-label versions (both endowed with the DM label, indicating bio-dynamic production) and the other product versions, which signal less strict organic standards. Given that most subjects did not know the DM label, the lack of statistically significant differences in ASRWTP between the versions with only the EU and those which additionally include the BL label might provide insights. Both labels are familiar to many

participants and relevant for purchase decisions, while a considerable number of subjects evaluates them as being non-uniform concerning the label category – which, in normative terms, could induce a WTP-premium. The absence of complementary effects is especially interesting, as there is evidence that in experiments, pro-social behavior is more accentuated than in the field (Benz and Meier, 2008). Thus, regarding the label combinations chosen in this study, it is questionable whether subjects are willing to pay more for a higher degree of sustainability in the grocery store if they are not even willing to do so in the laboratory.

It was shown above that according to the distribution test, we can reject the null hypothesis that bids are identical for unlabeled and labeled products, while we cannot reject the null hypothesis that bids are identical across labeled products, independent of the labels attached. In order to obtain a precise view of the drivers of the effects, Tobit models (Tobin, 1958) with Random Effects (RE) were tested.⁸ As we collect 16 WTP-values from each individual in a sequence, our product-sequence completely captures the “time-effect”, i.e. we can identify products with experimental rounds. In addition to the models for the whole product-sample, we also tested product-group-specific Tobit models in order to analyze whether there are differences in the valuation of labels depending on the type of product, and hence, to find an answer to our Exploratory Research Question. We tested two models for the whole sample and for each product group respectively, so as to differentiate a rather general label effect from label combination effects. The models in the style of Loureiro et al. (2013) are:

- (a) “Label”, describing a model which includes a dummy for labeled or non-labeled products;

$$ASRWTP_{it} = \max(-1; \alpha' \cdot l_{it} + \beta' \cdot x_{it} + u_i + \varepsilon_{it}) \quad (2)$$

where $ASRWTP_{it}$ is the marginal WTP of the i th subject for product t , l_{it} is a dummy indicating the existence of at least one label on the t th product bid for by subject i , x_i is a vector including socio-demographic control variables, u_i is the subject-specific, time-constant error term and ε_{it} the overall error term.⁹

- (b) “Version”, including dummies for the respective product versions;

$$ASRWTP_{it} = \max(-1; \alpha' \cdot v_{it} + \beta' \cdot x_{it} + u_i + \varepsilon_{it}) \quad (3)$$

where the label-dummy l_{it} from the “Label” model is replaced by a vector including dummies for each product version v_{it} .

Definitions and summary statistics of the included independent variables are given in Table 8. Model results are presented in Table 9. Concerning the influence of demographics, knowledge and attitudes, all models agree that women and people who have a bachelor’s degree yield a significantly higher ASRWTP than their respective counterparts. Additionally, the dummy indicating whether participants generally consider organic labels as important (“Organic_Imp”) significantly and positively influences ASRWTP when the models are run based on the whole sample. Yet, knowledge of any label does not significantly influence ASRWTP. Differences in ASRWTP thus cannot be explained by a lack of familiarity with the labels.

Concerning our hypotheses, the “Label”-models show that, for the whole sample as well as all product groups except animal-products, adding one or more label(s) significantly increases ASRWTP. Thus, the impression that labeled products yield significantly higher ASRWTP compared to unlabeled products is substantiated for three of the four product groups, while this effect is strongest among plant-based products, followed by non-perishable products. However, for animal products, no significant label effect is present in our data. This at least is surprising, as WTP for animal products might generally be considered more sensitive to labels than WTP for any other product group.

⁸ In order to decide whether our data is suitable for a RE model, we conducted a Hausman test. The result was that our data does quality for RE models. Moreover, we also conducted standard Tobit models without RE. The results are comparable to those we report in the RE Tobit model.

⁹ The selection of the independent variables accounts for multicollinearity.

Table 8: Summary statistics for independent variables.^h

	Definition	Mean	Std. Dev.
Age	Age. Categorical variable with 4 levels: 1: <20; 2: 20-24; 3: 25-29; 4: ≥ 30 .	2.3612	0.6923
Female	Gender. Dummy-variable; equals 1 if subject is female.	0.3998	0.4900
Bachelor	Bachelor's degree. Dummy-variable; equals 1 if subjects has a Bachelor's degree.	0.0802	0.2716
Net_Income	Monthly household net income. Categorical variable with 3 levels: 1: max.1500 Euros; 2: 1500-2500 Euros; 3: more than 2500 Euros.	1.8206	0.8843
Cook_Freq	Cooking frequency. Dummy-variable; equals 1 if subject cooks more than 4 times a month, up to several times a day.	0.6696	0.4705
Shopp_Freq	Shopping frequency. Dummy-variable; equals 1 if subject shops for food at least once a month.	0.8607	0.3463
Organic	Organic buyer. Dummy-variable; equals 1 if subject buys organic food on a regular basis.	0.4511	0.4977
FairTrade	Fair trade buyer. Dummy-variable; equals 1 if subject buys fair trade food on a regular basis.	0.3959	0.4892
Organic_Imp	Organic label importance. Dummy-variable; equals 1 if subject generally regards organic labels as important.	0.6197	0.4856
FairTrade_Imp	Fair trade label importance. Dummy-variable; equals 1 if subject generally regards fair trade labels as important.	0.7317	0.4432
Know_EU	Know EU. Dummy-variable; equals 1 if subject is very familiar with the EU label (knows it well).	0.1056	0.3074
Know_GER	Know GER. Dummy-variable; equals 1 if subject is very familiar with the GER label.	0.7722	0.4195
Know_NL	Know NL. Dummy-variable; equals 1 if subject is very familiar with the NL label.	0.0108	0.1032
Know_BL	Know BL. Dummy-variable; equals 1 if subject is very familiar with the BL label.	0.1730	0.3784
Know_DM	Know DM. Dummy-variable; equals 1 if subject is very familiar with the DM label.	0.0440	0.2051
Know_FT	Know FT. Dummy-variable; equals 1 if subject is very familiar with the FT label.	0.4746	0.4995
Label	Label. Dummy variable; equals 1 if a product is equipped with at least one label.	0.8822	0.3224
Version	Product version v . Categorical variable with 8 levels corresponding to the product versions.	3.5044	2.2835

^h Total number of subjects: 191. Calculations of means and standard deviations are based on 2046 observations of ASRWTP. The variables "KnowBL" and "FTLabelImp" are not included in the RE Tobit models due to multicollinearity.

Considering the "Version"-models, by tendency, we find further support for this result. In this model, for the whole sample as well as for plant-based products, all version-dummies for $v \geq 2$ yield significantly higher ASRWTP compared to the 0-label version. However, within the group of animal-products, no label combination increases ASRWTP significantly, compared to a version without any labels. Moreover, for non-perishable and perishable products, three of the seven (versions 3, 6, 7) respectively two out of seven (versions 2, 5) label combinations yield a significant WTP-premium. This result cannot be explained by the hypothesis of magnitude variation raised by normative theory, as it does not provide any plausible explanation for the lack of significance of, for example, the dummy for version 3 of perishable products. Why should there be a higher premium for version 2 than for version 3, if version 3 includes the same labels as version 2 and even one additional label?

Accordingly, the result could be classified as an anomaly in the valuation of food products. Following behavioral economic concepts, there are three explanations of the observed pattern. The phenomenon may be due to information-overload. However, the premiums attached for version 7 among non-perishable goods, for example, cannot be explained with this reasoning, as the information load is considerably higher for this version, compared to versions 2 and 4, which do not yield significant premiums.

Another explanation may be that there is a lack of environmental awareness, emotional involvement or locus of control on the consumer side, thus indicating the existence of attitude-behavior gaps. If this were true, it is questionable whether there would be any premiums for labels at all.

A third reasoning behind the inelasticity of WTP to the number of labels might be that subjects' valuations could be prone to an embedding effect. This latter explanation seems likely, as premiums do not vary systematically and significantly between label combinations, and as knowledge of the labels does not significantly influence bids, indicating that it does not matter whether people are familiar with the labels or not. Rather, it seems as if people generally make use of a "label-heuristic", indicating that labels are good and valuable, but which does not allow for a differentiated evaluation of specific labels.

Table 9: Random Effects Tobit models.ⁱ

	Whole Sample		Perishable		Non-Perishable		Animal		Plant-Based	
	Label	Version	Label	Version	Label	Version	Label	Version	Label	Version
Age	0.0814 (0.0543)	0.0791 (0.0555)	0.0458 (0.0712)	0.0451 (0.0749)	0.0356 (0.0721)	0.0404 (0.0767)	0.0234 (0.0661)	0.0235 (0.0711)	0.0479 (0.0698)	0.0476 (0.0727)
Female	0.1704** (0.0781)	0.1655** (0.0783)	0.1743* (0.1004)	0.1749* (0.1009)	0.3134*** (0.0983)	0.3103*** (0.1042)	0.2658*** (0.0921)	0.2669*** (0.0983)	0.2088* (0.1076)	0.2089* (0.1102)
Bachelor	0.5265*** (0.1058)	0.5242*** (0.1103)	0.7815*** (0.1330)	0.7865*** (0.1397)	0.2710* (0.1604)	0.2613 (0.1687)	0.5386*** (0.1462)	0.5393*** (0.1547)	0.5659*** (0.1637)	0.5471*** (0.1714)
Cook_Freq	-0.0429 (0.0788)	-0.0426 (0.0814)	-0.0194 (0.1033)	-0.0196 (0.1062)	-0.0089 (0.0942)	-0.0019 (0.0985)	-0.0109 (0.0937)	-0.0062 (0.0986)	0.0015 (0.1034)	0.0068 (0.1065)
Net_Income	-0.0408 (0.0446)	-0.0421 (0.0457)	-0.0210 (0.0554)	-0.0222 (0.0564)	-0.0500 (0.0504)	-0.0504 (0.0539)	-0.0277 (0.0491)	-0.0301 (0.0512)	-0.0223 (0.0583)	-0.0178 (0.0604)
Shopp_Freq	0.1087 (0.0900)	0.1056 (0.0905)	0.0884 (0.1221)	0.0868 (0.1266)	0.0118 (0.1208)	0.0061 (0.1269)	0.0023 (0.1196)	0.0060 (0.1253)	0.0989 (0.1230)	0.0873 (0.1272)
Organic	-0.0509 (0.0735)	-0.0545 (0.0754)	-0.0544 (0.1043)	-0.0545 (0.1074)	0.0955 (0.0978)	0.0958 (0.0995)	0.0425 (0.0961)	0.0383 (0.1037)	-0.0225 (0.1133)	-0.0283 (0.1176)
FairTrade	-0.0714 (0.0774)	-0.0723 (0.0787)	-0.0279 (0.1059)	-0.0288 (0.1101)	-0.1097 (0.1025)	-0.1094 (0.1063)	-0.0569 (0.0989)	-0.0580 (0.1037)	-0.0641 (0.1068)	-0.0615 (0.1177)
Organic_Imp	0.2036** (0.0822)	0.2057** (0.0837)	0.1990 (0.1266)	0.2012 (0.1283)	0.0962 (0.1192)	0.0905 (0.1230)	0.1257 (0.1177)	0.1221 (0.1217)	0.1669 (0.1418)	0.1677 (0.1458)
Know_EU	-0.0969 (0.1439)	-0.0979 (0.1498)	-0.1641 (0.1729)	-0.1644 (0.1765)	-0.0367 (0.1749)	-0.0353 (0.1575)	-0.1307 (0.1529)	-0.1302 (0.1575)	-0.1147 (0.1941)	-0.1139 (0.2002)
Know_GER	0.1257 (0.0765)	0.1253 (0.0800)	0.0735 (0.0979)	0.0746 (0.0992)	0.1008 (0.1196)	0.1006 (0.1223)	0.0256 (0.1081)	0.0273 (0.1134)	0.1441 (0.1120)	0.1326 (0.1165)
Know_NL¹⁰	0.0080 (1.2103)	0.0561 (0.7121)	-0.2585 (2.3024)	-0.1965 (2.2505)	-0.1410 (2.1485)	-0.1419 (2.1267)	-0.2038 (14.4406)	-0.1725 (5.2110)	-0.2712 (2.6866)	-0.3099 (2.1236)
Know_DM	0.0247 (0.2728)	0.0304 (0.2708)	0.0087 (0.4614)	0.0047 (0.4671)	0.0669 (0.3044)	0.0641 (0.3104)	0.0061 (0.3625)	0.0130 (0.3656)	0.0467 (0.2949)	0.0447 (0.3142)
Know_FT	-0.1143 (0.0806)	-0.1165 (0.0825)	-0.0779 (0.1021)	-0.0798 (0.1030)	-0.0341 (0.1128)	-0.0326 (0.1218)	-0.0010 (0.1063)	-0.0012 (0.1124)	-0.0736 (0.1121)	-0.0643 (0.1176)
v = 1	--	0.1013 (0.0759)	--	0.1545 (0.1181)	--	0.0696 (0.1148)	--	0.0648 (0.1016)	--	0.1461 (0.1336)
v = 2	--	0.1568** (0.0614)	--	0.1824* (0.0987)	--	0.1332 (0.1020)	--	0.1032 (0.0821)	--	0.2091* (0.1095)
v = 3	--	0.1456** (0.0655)	--	0.0961 (0.1233)	--	0.1888* (0.1017)	--	0.1081 (0.0824)	--	0.1979* (0.1193)
v = 4	--	0.1502** (0.0708)	--	0.1348 (0.0971)	--	0.1625 (0.1351)	--	0.0717 (0.0928)	--	0.2262* (0.1321)
v = 5	--	0.1640** (0.0697)	--	0.1715* (0.1017)	--	0.1785 (0.1171)	--	-0.0491 (0.1022)	--	0.3611*** (0.1250)
v = 6	--	0.1448** (0.0639)	--	0.1140 (0.1053)	--	0.2033* (0.1051)	--	0.0761 (0.0818)	--	0.2187* (0.1268)
v = 7	--	0.1927*** (0.0656)	--	0.1646 (0.1204)	--	0.2544** (0.0995)	--	0.1248 (0.0785)	--	0.2440* (0.1336)
Label	0.1507*** (0.0495)	--	0.1461* (0.0851)	--	0.1698** (0.0789)	--	0.0707 (0.0608)	--	0.2281** (0.0951)	--
Observations	2046	2046	1116	1116	930	930	921	921	1125	1125
Log Likelihood	-2163.3550	-2161.6340	-1140.1930	-1137.9960	-1068.1000	-1065.6740	-959.3918	-955.6944	-1260.6170	-1255.5560

ⁱ Standard errors are reported in parenthesis. The dependent variable is ASRWTP as defined in the text, i.e. it depicts the marginal WTP for a certain product version, compared to conventional products in the supermarket.

*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

¹⁰ Know_NL has only few observations =1 and standard errors are high. Thus, we replicated all models without Know_NL and the results are comparable to those reported here. This holds if Know_BL is included instead in the models (while Know_BL itself has no significant influence on ASRWTP in any of the models).

7 Conclusion

Concerning the question of whether sustainability labels on groceries increase WTP compared to unlabeled products, this study provides answers in the affirmative. As the experimental results show that labels increase median WTP by up to 14.45%, they serve as instruments for signaling information which customers value positively. However, labels appear to be only imperfect signals, mobilizing rather broad heuristics instead of supporting differentiated decision-making. Accordingly, the specific type of sustainability information conveyed by different labels appears to be of minor importance, as respondents are insensitive to the number of labels, even if additional labels add new dimensions of label criteria and increase the level of certification strictness. It is noteworthy that this effect is independent of knowledge of the labels and cannot be explained by a lack of ability to discriminate between the labels. Accordingly, these results cannot be interpreted solely as substitution effects, but rather as anomalies in consumer valuation and are thus in line with behavioral decision theories. Likely, our participants are prone to an embedding effect, suggesting that they do not obtain a proportional utility increase from another label, but rather evaluate sustainability as an “all or nothing” concept.

For future work, it would be interesting and useful to substantiate our findings and to extend the sample. Moreover, even though we aimed at providing a broad picture of the product- and label-landscape relevant for our participants, other labels, other products and other product brands might induce somewhat different results. The extrapolation of our findings to other populations, products and labels should therefore be treated with caution. Furthermore, our experimental design does not control for product-ordering effects by itself, and our RE Tobit model might not fully correct them. Finally, with regard to the repeated-measures design and the use of sustainability labels, learning-effects and social-desirability bias cannot be ruled out, although the crossover design includes a variety of mechanisms to reduce these problems.

References

- Alfnes, F., Rickertsen, K., 2011. Nonmarket Valuation: Experimental Methods, in J. L. Lusk, J. Roosen and J. Shogren. The Oxford Handbook of The Economics of Food Consumption and Policy (Oxford Handbooks in Economics). Oxford University Press, Oxford, 215–242.
- Bauer, H. H., Heinrich, D., Schäfer, D. B., 2013. The Effects of Organic Labels on Global, Local, and Private Brands. More than Substance? Journal of Business Research. 8(66), 1035–1043.
- Becker, G. M., Degroot, M. H., Marschak, J., 1963. Stochastic Models of Choice Behavior. Behavioral Science. 8(1), 41–55.
- Becker, G. M., Degroot, M. H., Marschak, J., 1964. Measuring Utility by a Single-Response Sequential Method. Behavioral Science. 9(3), 226–232.
- Benz, M., Meier, S., 2008. Do People Behave in Experiments as in the Field? Evidence from Donations. Exper Econ. 11(3), 268–281.
- Bernard, J. C., Bernard, D. J., 2009. What is it About Organic Milk? An Experimental Analysis. American Journal of Agricultural Economics. 91(3), 826–836.
- Bernard, J. C., He, N., 2010. Confounded by The Field: Bidding in Food Auctions when Field Prices are Increasing. Agricultural and Resource Economics Review. 39(2), 275–287.
- Bernard, J. C., He, N., 2011. Differences in WTP and Consumer Demand for Organic and Non-GM Fresh and Processed Foods. Agricultural and Resource Economics Review. 40(2), 218–232.
- Bernard, J. C., Zhang, C., Gifford, K., 2006. An Experimental Investigation of Consumer Willingness to Pay for Non-GM Foods when an Organic Option is Present. Agricultural and Resource Economics Review. 35(2), 374–385.
- Bohm, P., Lindén, J., Sonnegard, J., 1997. Eliciting Reservation Prices: The Becker DeGroot Marschak Mechanisms versus Markets. The Economic Journal. 107, 1079–1089.
- Bougherara, D., Grolleau, G., Mzoughi, N., 2007. Is More Information Always Better? An Analysis Applied to Information-Based Policies for Environmental Protection. International Journal of Sustainable Development. 10(3), 197–213.
- Brown, A. L., Cohen, G., 2015. Does Anonymity Affect the Willingness to Accept and Willingness to Pay Gap? A Generalization of Plott and Zeiler. Exp Econ. 18(2), 173–184.
- Brunner, E., Munzel, U., 2000. The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. Biometrical Journal. 42(1), 17–25.
- Chang, J. B., Lusk, J. L., Norwood, F. B., 2009. How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behavior? American Journal of Agricultural Economics. 91(2), 518–534.
- Clamor, T., 2010. Determinanten für nachhaltigen Konsum – Eine empirische Analyse für Deutschland. Soziale, ökologische und ökonomische Dimensionen eines nachhaltigen Energiekonsums in Wohngebäuden.
- Costanigro, M., Kroll, S., Thilmany, D., Bunning, M., 2014. Is it Love for Local / Organic or Hate for Conventional? Asymmetric Effects of Information and Taste on Label Preferences in an Experimental Auction. Food Quality and Preference. 31, 94–105.
- Cummings, R. G., Harrison, G. W., Rutström, E. E., 1995. Homegrown Values and Hypothetical Surveys. Is the Dichotomous Choice Approach Incentive-Compatible? American Economic Review. 85, 260–266.
- Degeratu, A. M., Rangaswamy, A., Wu, J., 2000. Consumer Choice Behavior in Online and Traditional Supermarkets: The Effects of Brand Name, Price, and Other Search Attributes. International Journal of Research in Marketing. 17(1), 55–78.
- Devinney, T. M., Auger, P., Eckhardt, G. M., 2010. *The Myth of the Ethical Consumer*. Cambridge Univ. Press, Cambridge u.a.
- Dufeu, I., Ferrandi, J.-M., Gabriel, P., Le Gall-Ely, M., 2014. Socio-Environmental Multi-Labeling and Consumer Willingness to Pay. Recherche et Applications en Marketing (English Edition). 29(3), 35–56.
- Einhorn, H. J., Hogarth, R. M., 1981. Behavioral Decision Theory. Processes of Judgment and Choice. Journal of Accounting Research. 32(1), 53–88.
- Eppler, M. J., Mengis, J., 2008. The Concept of Information Overload. A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. Kommunikationsmanagement im Wandel, 271–305.
- European Union, 2015. Sustainable Development, available at <http://ec.europa.eu/environment/eussd/>.
- Fischbacher, U., 2007. Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. Exp Econ. 10(2), 171–178.
- Fischhoff, B., Jacobs Quadrel, M., Kamlet, M., Loewenstein, G., Dawes, R., Fischbeck, P., Klepper, S., Leland, J., Strohm, P., 1993. Embedding Effects. Stimulus Representation and Response Mode. J Risk Uncertainty. 6(3), 211–234.
- Gracia, A., Barreiro-Hurlé, J., López-Galán, B., 2014. Are Local and Organic Claims Complements or Substitutes? A Consumer Preferences Study for Eggs. J Agric Econ. 65(1), 49–67.
- Guala, F., 2000. Artefacts in Experimental Economics: Preference Reversals and the Becker-DeGroot-Marschak Mechanism. Economics and Philosophy. 16(1), 47–75.
- Harper, G. C., Makatouni, A., 2002. Consumer Perception of Organic Food Production and Farm Animal Welfare. British Food Journal. 104(3/4/5), 287–299.
- Irwin, J. R., Spira, J. S., 1997. Anomalies in the Values for Consumer Goods with Environmental Attributes. Journal of Consumer Psychology. 6(4), 339–363.

- Jongmans, E., Jolibert, A., Irwin, J., 2014. Is More Always Better? Counter-intuitive Effects in Consumer Assessment of Environmental Product Attributes. *Recherche et Applications en Marketing (English Edition)*. 29(3), 10–34.
- Kahneman, D., Knetsch, J. L., 1992. Valuing Public Goods: The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management*. 22(1), 57–70.
- Kehlbacher, A., Bennett, R., Balcombe, K., 2012. Measuring the Consumer Benefits of Improving Farm Animal Welfare to Inform Welfare Labelling. *Food Policy*. 37(6), 627–633.
- Kollmuss, A., Agyeman, J., 2002. Mind the Gap: Why Do People Act Environmentally and What are the Barriers to Pro-Environmental Behaviour? *Environmental Education Research*. 8(3), 239–260.
- Larceneux, F., Benoit-Moreau, F., Renaudin, V., 2012. Why Might Organic Labels Fail to Influence Consumer Choices? Marginal Labelling and Brand Equity Effects. *J Consum Policy*. 35(1), 85–104.
- Loureiro, M. L., Gracia, A., Nayga, R. M., 2013. Do experimental auction estimates pass the scope test? *Journal of Economic Psychology*. 37, 7–17.
- Loureiro, M. L., Lotade, J., 2005. Do Fair Trade and Eco-Labels in Coffee Wake Up the Consumer Conscience? *Ecological Economics*. 53(1), 129–138.
- Loureiro, M. L., McCluskey, J. J., Mittelhammer, R. C., 2001. Assessing Consumer Preferences for Organic, Eco-Labeled, and Regular Apples. *Journal of Agricultural and Resource Economics*. 26(2), 404–416.
- Loureiro, M. L., McCluskey, J. J., Mittelhammer, R. C., 2002. Will Consumers Pay a Premium for Ecolabeled Apples? *Journal of Consumer Affairs*. 36(2), 203–219.
- Luchs, M. G., Walker Naylor, R., Irwin, J. R., Raghunathan, R., 2010. The Sustainability Liability: Potential Negative Effects of Ethicality on Product Preference. *Journal of Marketing*. 74(5), 18–31.
- Lusk, J. L., Feldkamp, T., Schroeder, T. C., 2004. Experimental Auction Procedure. Impact on Valuation of Quality Differentiated Goods. *American Journal of Agricultural Economics*. 86(2), 389–405.
- Muller, L., Ruffieux, B., 2011. Do Price-Tags Influence Consumers' Willingness to Pay? On the External Validity of Using Auctions for Measuring Value. *Exp Econ*. 14(2), 181–202.
- Onozaka, Y., Mc Fadden, D. T., 2011. Does Local Labeling Complement or Compete with Other Sustainable Labels? A Conjoint Analysis of Direct and Joint Values for Fresh Produce Claim. *American Journal of Agricultural Economics*. 93(3), 693–706.
- Pedersen, E. R., Neergaard, P., 2006. Caveat Emptor - Let the Buyer Beware! Environmental Labelling and the Limitations of 'Green' Consumerism. *Bus. Strat. Env.* 15(1), 15–29.
- Pelsmacker, P. de, Driesen, L., Rayp, G., 2005. Do Consumers Care about Ethics? Willingness to Pay for Fair-Trade Coffee. *The Journal of Consumer Affairs*. 39(2), 363–385.
- Sénéchal, S., Georges, L., Pernin, J. L., 2014. Alliances Between Corporate and Fair Trade Brands: Examining The Antecedents of Overall Evaluation of the Co-branded Product. *J Bus Ethics*. 124(3), 365–381.
- Shi, L., House, L. A., Gao, Z., 2013. Impact of Purchase Intentions on Full and Partial Bids in BDM Auctions: Willingness-to-Pay for Organic and Local Blueberries. *J Agric Econ*. 64(3), 707–718.
- Sirieux, L., Delanchy, M., Remaud, H., Zepeda, L., Gurviez, P., 2013. Consumers' Perceptions of Individual and Combined Sustainable Food Labels: A UK Pilot Investigation. *International Journal of Consumer Studies*. 37(2), 143–151.
- Sirieux, L., Tagbata, D., 2008. Measuring Consumer's Willingness to Pay for Organic and Fair Trade Products. *International Journal of Consumer Studies*. 32(5), 479–490.
- Sörqvist, P., Hedblom, D., Holmgren, M., Haga, A., Langeborg, L., Nösth, A., Kågström, J., 2013. Who Needs Cream and Sugar when there is Eco-Labeling? Taste and Willingness to Pay for "Eco-Friendly" Coffee. *PLoS ONE*. 8(12).
- Tobin, J., 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*. 26(1), 24–36.
- van Huylenbroek, G., Aertsens, J., Verbeke, W., Mondelaers, K., 2009. Personal Determinants of Organic Food Consumption: A Review. *British Food Journal*. 111(10), 1140–1167.
- Vermeir, I., Verbeke, W., 2006. Sustainable Food Consumption: Exploring The Consumer "Attitude – Behavioral Intention" Gap. *J Agric Environ Ethics*. 19(2), 169–194.
- Vickrey, W., 1961. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*. 16, 8–37.
- Wertenbroch, K., Skiera, B., 2002. Measuring Consumers' Willingness to Pay at the Point of Purchase. *Journal of Marketing Research*. 39(2), 228–241.
- Zizzo, D. J., 2010. Experimenter Demand Effects in Economic Experiments. *Exp Econ*. 13(1), 75–98.