International Food and Agribusiness Management Review
Volume 19 Issue 3, 2016

Can a Multi-Level Label do Better than a Binary Label for Animal Welfare? A PLS-Analysis of Consumer Satisfaction

Ramona Weinrich [Φa] and Achim Spiller [b]

[a]*Post-Doctoral Researcher, Georg-August-University Goettingen, Department of Agricultural Economics and Rural Development Management in Agribusiness, Centre of Biodiversity and Sustainable Land Use, Platz der Göttinger Sieben 5, 37073 Goettingen, Germany*

[b]*Professor, Georg-August-University Goettingen, Department of Agricultural Economics and Rural Development Marketing for Food and Agricultural Products, Platz der Göttinger Sieben 5, 37073 Goettingen, Germany*

Abstract

Labeling is an important marketing tool for food producers and retailers. With growing product heterogeneity, labels can help consumers differentiate higher quality products on supermarket shelves. Currently, most labels are of a binary nature—meaning a product either has certain characteristics or not—although there is a larger product heterogeneity in the food market than just two standards. A multi-level label might be a solution to addressing this problem. The objective of this article is to investigate if influences on consumer satisfaction with ethical food labeling systems differ between a binary and a multi-level labeling system. A consumer survey was carried out in Germany (n = 1,538) comparing the two types of labels with a split-sample approach. The influence of five factors (comprehensibility, involvement, time pressure, trust and use), derived from the literature, were analyzed in a structural equation model. All these factors influence satisfaction with labeling. Furthermore, differences between the two labeling systems were detected. This article delivers important results for food producers and policy makers. The group comparison indicated that trust as a precondition is more necessary for a binary label whereas time pressure factors reduce satisfaction with multi-level labeling.

Keywords: labeling, consumer research, animal welfare, structural equation modeling,

[Φ]Corresponding author: Tel: +49 551 39 19534
Email: R. Weinrich: ramona.weinrich@agr.uni-goettingen.de
A. Spiller: a.spiller@agr.uni-goettingen.de

Introduction

Labeling is an important marketing tool for food producers and retailers. Although van der Merwe et al. (2010) found that habitual purchasing is more important than detailed product information on packages; it is known that with growing product heterogeneity, labels can help consumers differentiate higher quality products on supermarket shelves. Additionally, labeling is the most important means of reducing information asymmetry (Akerlof 1970; Darby and Karni 1973; Nelson 1970). Compared to consumers, producers usually have more information about their products and the underlying production processes. This is especially true for credence attributes, which cannot be verified by consumers when purchasing a product (Darby and Karni 1973). For example, a credence attribute can be an animal-welfare-related production method as seen in the practice of limiting floor space per animal for the purpose of pig fattening. Therefore, labels can help consumers who are looking for special products by turning credence attributes into search goods (Caswell and Mojduszka 1996; Caswell and Padberg 1992; Jahn et al. 2005). Thus, labels are frequently used for especially ethical attributes, which are usually credence attributes.

Ethical consumerism is on the rise: in Germany for example, 30% of consumers actively consider two or more ethical issues when making purchasing decisions (UK: 29%, France: 24%, Spain: 9%); a further 31% consider some ethical issues but not habitually (UK: 28%, France: 26%, Spain: 18%) (IGD.com 2010). In the food sector, the ethical consumerism trend is expressed by a wide range of ethical labels. The different labels identify products to be produced to certain ethical production standards, such as positive frames (e.g. pasture-raised beef), negative frames (e.g. GM free) or value neutral frames that stand for a certain production method (e.g. organic farming) (Schröder and McEachern 2004). However, although consumers' concerns about ethical problems have been well documented, skepticism about ethical labeling is still prevalent (Hoek et al. 2013). Currently, most ethical food labels are of a binary nature, meaning that a product either has a certain characteristic or does not. Hence, by means of a binary label, a complex production or process method such as animal welfare is reduced to one dimension. The product is labeled animal welfare or not. Nevertheless, many food characteristics that are gaining in importance nowadays require more detailed labels to display their quality information accurately such as complex ethical processes or product standards. A multi-level label might be an appropriate solution to address this problem as it can show consumers that there is not only "good and bad" in the food market but product heterogeneity. In this way, a differentiated labeling strategy might defuse the discussion about "good and bad" foods, the resulting "food wars" and the public debate in the mass media which has been fueled by deep controversies between a productionist versus an ecologically integrated paradigm (Lang and Heasman 2004).

At the moment, there are only a few examples of multi-level labels for ethical food. One is an animal welfare label called *Beter Leven* in the Netherlands with one to three stars. Another example is the 5-Step® Animal Welfare Rating standards which can be found in Whole Foods Market chain in the US. There is also a NOP labeling system in the US which uses four levels showing the gradations of organic content: *100% Organic; Organic* (95% or more organic ingredients); *Made with organic ingredients* (at least 70% organic ingredients); and *Specific organic ingredients* (USDA–AMS 2012).

Yet there are already two world-renowned multi-level labeling systems in the non-food sector: energy class labeling (e.g. for electronic products, buildings, cars or TVs) and hotel classification (stars or diamonds). It is not clear, however, whether a multi-level label might be a long-term solution for ethical food labels as well. This study was therefore primarily aimed at testing a multi-level label in the food sector. A preliminary study has shown that when a two-level labeling scheme is explained to consumers, willingness to pay rises with the label standard (Weinrich, Franz, and Spiller *Forthcoming*). Furthermore, Weinrich and Spiller (2016) showed that a multi-level animal welfare label can achieve higher market shares, increase animal welfare levels and result in higher sales.

In general, the following study amends the growing body of literature about food labeling with a special focus on an unexplored part of the debate. The innovation of this article is a measurement of the level of satisfaction and its drivers comparing a binary and a multi-level label in a split sampling design.

Multi-Level Labeling Schemes

In contrast to the food sector there are established multi-level labeling schemes for hotels and energy. With revealed preference methods for these sectors it is therefore possible to calculate willingness to pay for different quality levels, either with regression analyses or hedonic pricing models.

For the hotel market, Espinet et al. (2003) showed a significant rise in prices from a 3-star to a 4-star hotel. Abrate et al. (2011) confirmed that a higher classified hotel has higher prices. Furthermore, Zhang et al. (2011) calculated that room prices in New York are 35.8% higher with each incremental star. However, on the contrary, Núñez-Serrano et al. (2014) found that the different classification systems in the hotel industry in Spain could lead to a loss of credibility. Su and Sun's (2007) results were similar—hotel rating systems should be updated and incorporate consumers' viewpoints in order to gain more acceptance.

In terms of energy efficiency, Galarraga et al. (2014) showed for the car market that consumers buy relatively rationally. They determine their price structure by including the respective fuel saving which the brands promise. A and B labeled cars are sold with surcharges between 3.0 and 5.9% more than similar but less energy-efficient cars. Consumers even pay more for the car than the value of fuel saving. Also for cars, Alberini et al. (2016) calculated that when a car is A-labeled in energy efficiency, there is a surplus of 6–11% on the selling price. Furthermore, Galarraga et al. (2011) found that for energy efficient labeled dishwashers 15.6% of the price is due to the label. For buildings, Eichholtz et al. (2010) revealed that prices for green rated properties are around 3% higher per square foot than for otherwise identical buildings.

However, the transfer of such results to the food market must be limited and careful. Hotels belong to the service sector, not to the fast moving consumer-good sector. Energy efficiency is a unidimensional sustainability characteristic referring only to energy which is much easier to understand in comparison to animal welfare, which is a multidimensional construct that includes, for example, space per animal, use of medicines or animal behavior. In addition, the incentive to

buy energy-efficient labeled products are most likely about saving money in the long run which is not the case when buying animal welfare labeled products with altruistic motives.

For the food market, there has been a variety of studies with a focus on nutrition labeling. Nutrition labeling can be binary (e.g. "Healthier Choice Tick" or "Smileys"), but can also be multi-level ("Traffic light" [TL; rating saturated fat, sugar, and sodium per serving], "Wheel of Health" and "GDA [Guideline Daily Amount] scores") Andrews et al. (2011) compared a binary nutrition label indicating a healthier choice (Smart Choices Program) to a TL labeling scheme in their study. Although the participants stated that they preferred the simpler label, the binary label could lead to positive (and potentially misleading) nutrient evaluations and product-healthiness perceptions when compared to a multi-level labeling system (ibid.). These findings are supported by Kelly et al. (2009), who found that consumers could identify healthier food with the TL system five times better than with the GDA system and three times better than with a colored GDA system. This work has been complemented by Roberto et al. (2012), who conducted an analysis that used nutrient quizzes. They found that participants achieved the best overall performance with the TL+ (additionally rating fiber and protein per serving) system when compared to no label, TL and GDA labeling systems.

These results from the literature indicate that a multi-level label might have clear benefits compared to a binary labeling system. However, these findings cannot be transferred to ethical labeling directly because, while nutrient and nutrition labeling refers to physical aspects, ethic labeling has different, often motivational, underlying consumption reasons. Additionally, ethical aspects often do not affect product attributes themselves but process attributes. However, Fisher et al. (2013) have suggested the use of a multi-level ecolabel based on theoretical calculations.

Objectives and Conceptual Framework

For the success of a food label which aims to reduce information asymmetry, it is important to meet the information needs of the target group (Verbeke 2005). If the information provided by a label is sufficient, satisfaction with the label and therefore with the product rises and repeated purchase is probable (Grunert 2002; Scott and Worsley 1994). The given amount of information differs between a binary and a multi-level label. Thus, the objective of this article is to investigate whether influences on consumer satisfaction with ethical food labeling systems differ between a binary and a multi-level labeling system. In this way, it can be analyzed whether an ethical multi-level label can be considered superior to a binary label and so deliver important results for both food producers and policy makers.

Our study uses an experimental animal welfare label to test the viability of using a multi-level label in the food market due to the fact that animal welfare is representative for ethical food labels as e.g. discussed by Schröder and McEachern (2004). Furthermore, animal welfare is discussed intensely in many industrialized countries (e.g. Clark et al. 2016). Since no well-known multi-level label exists for animal welfare in the German market, a potential bias of the study due to images of existing labels could be prevented.

In detail, this study was aimed at identifying the potential problems and advantages of a multi-level label in comparison to a binary label. This was achieved by using a split-sample design to

differentiate between both approaches. Again, to prevent potential bias, the binary as well as the multi-level label were designed specifically for the study and do not currently exist in the market. The two facets of this study were undertaken as they will contribute to the consumer research that needs to be conducted prior to introducing a multi-level label for ethical food products.

In order to assess consumers' satisfaction with a labeling scheme, influencing factors had to be identified. From the literature, the following five factors influencing labeling satisfaction were determined:

1-Comprehensibility

Comprehensibility is an essential factor for the distribution and success of a food label. If a label is not understood by consumers, it does not facilitate the weekly shopping. It adds no value so consumers would almost certainly be dissatisfied with such a labeling scheme. Aarset et al. (2004) found that in Europe, consumer knowledge about what organic and respective ecolabels actually mean is not sound. This was confirmed by Buxel and Schulz (2010) in a consumer survey with German consumers—less than 50% of the participants indicated that they knew all about food labels and what particular labels actually revealed. Consumers did not possess any specific information about most labels. An explanation for these results might be found in the so-called 'information overload,' whereby consumers are confronted with a mass of information in everyday life—not only concerning food (Kolodinsky 2012; Kroeber-Riel and Esch 2004; van Kleef et al. 2008; Verbeke 2005). Due to limited cognitive capacity not all information provided on food packages, especially in textual form, can be processed – viz. limited willingness and opportunity to process information on food packages (Verbeke 2005). This is supported by the results from Schulte-Mecklenbeck et al. (2013), who confirmed that very little time has been spent on the decision-making process in order to understand and evaluate all relevant factors; instead, heuristics are applied. Furthermore, these authors found that visual stimuli gained more attention than numeric illustrations. Caputo et al. (2013) supported these findings regarding climate-friendly labeling. They found that environmental information related to time and distance is preferred by consumers in comparison to abstract numbers in the form of a carbon footprint.

From the above literature, it is inducible that the influence of comprehensibility on satisfaction with a label system is not predictable. A tiered presentation of different standards in food processing and product standards might either increase the comprehensibility of a multi-level label by transparently showing different standards or conversely, it might decrease comprehensibility, despite its transparency, due to complexity. The first hypothesis is derived as a result of the literature analysis:

H_1: Comprehensibility of food labels influences consumer satisfaction with an ethical food labeling system.

2–Involvement

Usually, consumer involvement in everyday products in the food sector is rather low as purchases are characterized as habitual in developed countries (Aertsens et al. 2009). However,

knowledge may help increase food involvement (ibid.). Silayoi and Speece (2004) confirmed that visual packaging plays a major role in low involvement decisions. Here, a multi-level label can provide more information in a visual form than is possible with a binary-labeling system. However, if consumers are characterized by higher involvement, more information is usually sought (ibid., Spiller 2010). This has been confirmed by Espejel et al. (2009), who found that the influence of quality attributes on consumers' satisfaction with a labeling system depends on whether the consumers have a high or a low involvement. Consequently, the second hypothesis is:

H_2: Involvement influences consumer satisfaction with an ethical food labeling system.

3–Time Pressure

As shown in the previous section concerning involvement, most food decision-making processes are characterized by low involvement. In accordance with this is the fact that the time consumers spend on shopping has decreased over the last decade (Feunekes et al. 2008). Even seventeen years ago, Warde (1999) confirmed that consumers were facing more time constraints in everyday life, which resulted in an increased consumption of convenience food. Likewise, Darian and Cohen (1995) emphasized in an earlier study the time constraints for purchasing and preparing food as well as food consumption itself. Moreover, the analyses of Schulte-Mecklenbeck et al. (2013) showed that decision-making in terms of food purchasing is based on the assumption of limited search and also due to time constraints in everyday life. In addition, in a study carried out by van der Merwe et al. (2010), participants stated that their disinterest in label information was due to time pressure. Thus, time pressure can also be a variable that influences label satisfaction. A consequence of this pressure might be less time spent reading and understanding food labeling. Consumers' satisfaction with labeling could then be influenced in one of two ways: if consumers feel under time pressure, they could feel confronted with too much information on a multi-level label or they might find the information easier to access compared with a binary label as they do not have to read additional information. In order to test these assumptions in empirical research, the third hypothesis is formulated:

H_3: Time pressure influences consumer satisfaction with an ethical food labeling system.

4–Trust

Trust is an important construct concerning food labeling as a label transfers credence attributes into a search good (Caswell and Mojduszka 1996; Caswell and Padberg 1992; Jahn et al. 2005). However, Buxel and Schulz (2010) found that trust in labels is very variable depending on the particular label. For example, the Demeter label, a label from a German organic farming association with higher standards than the legal minimum for organic products, had 90% agreement for trust. The German governmental organic label was only trusted by 68% of the respondents. These agreement rates from German consumers can still be considered high after Aertsens et al. (2009) detected in a literature review, an overall lack of trust in the certification process for organic food. Specifically, Krystallis and Chryssshoidis (2005) found that less than one-third of Greek consumers trusted retail outlets which distributed certified organic products, reflecting great distrust in those vendors of organic products.

This lack of trust could possibly be explained by the fact that differences between labels are not transparent for consumers. A multi-level label might gain more trust as product and process differences are more transparently presented. This might in turn effect satisfaction with the label. In an Australian consumer study conducted by Williams and Mummery (2012), the results showed that high proportions of consumers trusted the Heart Foundation Tick which signals healthy food. Logically, trust in a label is accompanied by a higher degree of satisfaction (ibid.).

In contrast, Fenko et al. (2016) modeled skepticism (a related concept) rather than trust in food labels. According to their results, skepticism (i.e. lack of trust) is a crucial factor that influences consumer responses to food. With respect to the present research question, therefore, it is possible that this would mean the more consumers trust a label, the more they are satisfied with that label. However, consumers who have higher generalized trust are no more likely to buy environmentally labeled products than those who have lower trust (Grebitus et al. 2015). Unlike this result, Roosen et al. (2015) showed that German and Canadian consumers' willingness to pay for new food characteristics increases when trust is higher. Thus, trust is essential when introducing innovative products which imply higher prices. Herrera and Blanco (2011) analyzed trust as a precursor for satisfaction. They also found that satisfaction influences willingness to buy and the purchase frequency. This was confirmed by Nocella et al. (2010, 2014). These authors ascertained that trust in a label is essential and that trust raises the willingness to pay a price premium for products with high animal welfare certification. Thus, due to this potentially higher willingness to pay, producers and retailers should also consider trust in their food labeling.

An analysis is required on whether a multi-level label system can gain more trust due to grading in its design and if the subsequent label satisfaction is higher compared to a binary label system. Hence, the fourth hypothesis is:

H_4: Trust in food labels influences consumer satisfaction with an ethical food labeling system.

5–Use

A label is only a helpful search attribute for consumers' decision making if it is used frequently. Thus, for a labeling system to succeed, it is crucial to examine whether satisfaction with a multi-level or a binary labeling system is influenced by use. However, this question has not been addressed in research as yet. Generally, despite limited label knowledge, consumers have a positive attitude towards food labeling. According to Buxel and Schulz (2010), 85% of German consumers think that a food label is a good thing, and 80% say that labels are useful. Besler et al. (2012) stated that about 75% of Turkish consumers report using food labels, with barriers for use being the lack of understanding of terms, symbols and values together with poor presentation of information. Both these sets of authors' results might be interpreted as being in favour of a multi-level labeling system as such a design could be understood more intuitively, would increase use, which in turn would raise consumer satisfaction. In contrast, other results from empirical research favour binary labeling systems. Although label information is used by nearly all consumers, they still prefer simplified labeling (Silayoi and Speece 2004). However, it may even be that the label's degree of comprehensibility is less important with respect to use, as Verbeke et al. (2012) proved that label use is closely connected with interest.

There is also still a percentage of consumers who do not use food labels at all. Such consumers rely instead on the availability of food, freshness, quality and the impression of the package design itself (van der Merwe et al. 2010). This would mean that use itself had no influence on consumer satisfaction with a labeling system. In order to analyze whether label use has an influence on satisfaction with a labeling system, the fifth hypothesis is proposed:

H_5: Food label use influences consumer satisfaction with an ethical food labeling system.

The five above-mentioned factors were modelled as latent variables influencing consumer satisfaction with a labeling system. In a consumer study, the participants were presented with either an ethical binary or a multi-level label. This split sampling approach serves as a moderator variable. Figure 1 shows the relationships between the latent variables used in our model. The underlying items are shown in Table A2 in the Appendix. Finally, a multi-group comparison between the binary and the multi-level split for all five hypotheses will be carried out in order to test for differences between the two labeling systems.

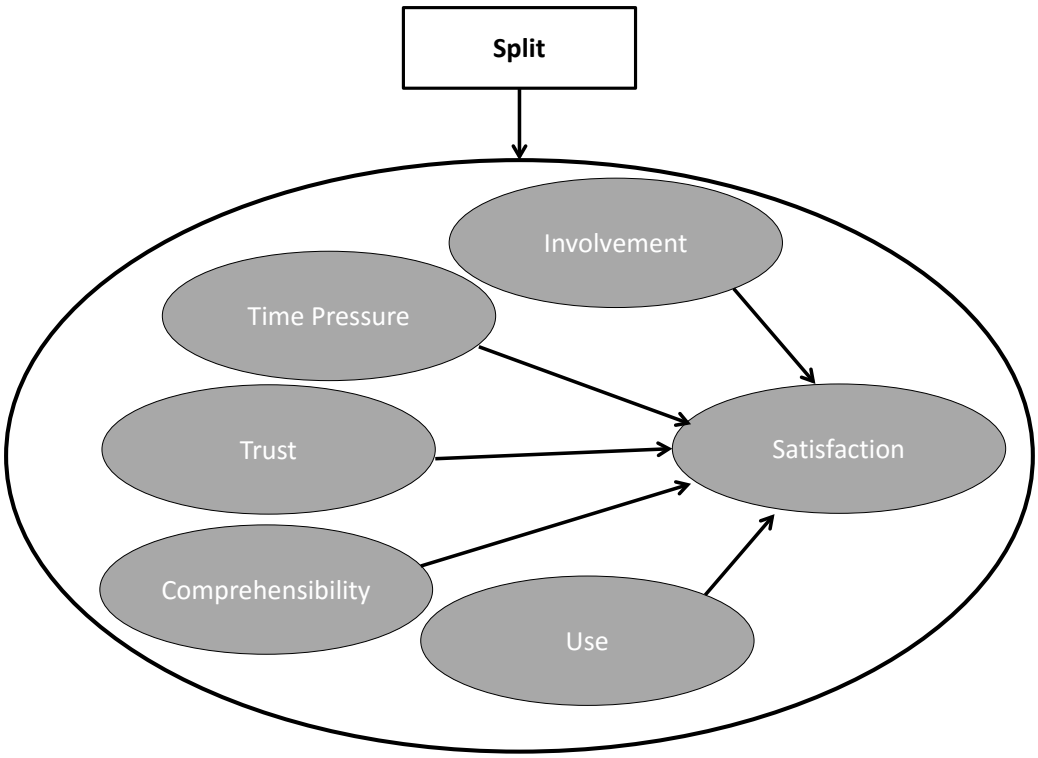

Figure 1. Moderator modeling framework
Source. Authors' illustration

Material and Methods

The data collection took place in 2014 by means of an online access panel. The sample size is 1,538 German consumers. Quotas were set for age, gender, education and income to ensure maximum representation from the German population.

Split-sampling design signifies that the participants were randomly allocated to one of two groups so that they either saw the binary or the multi-level label. The quotas mentioned above were also set within each sub-sample (n = 769).

The questionnaire had the following structure:

At the beginning of the survey, the participants were given information about and shown either an experimental binary or a multi-level label. The consumers received two examples of the standards of the label: floor area per animal and transport time requirements to the slaughterhouse (the minimum standards used are those legally required by the Federal Republic of Germany). These two standards were chosen as examples because transportation time has an influence on animal welfare (Vecerek et al. 2006) and is an important and straightforward aspect for consumers (Pouta et al. 2010). From a scientific perspective, stocking density is also an important factor for animal welfare (Bokkers et al. 2011; Talebi et al. 2014; Turnbull et al. 2005) and is considered to be essential information for animal welfare oriented customers (de Jonge and van Trijp 2013; Vanhonacker et al. 2009).

After this introduction, consumers had to make a choice between conventionally produced meat and the labeled product (binary split) or one of the labeled alternatives (multi-level split) and the label information was provided a second time[1]. Subsequently, there were questions about attitude towards the label. Further questions regarding quality aspects, labeling, and purchasing behavior were the same for both sub-samples.

The respondents scored their answers on five-point Likert scales or Likert-like scales. The Likert scales for the statements concerning time pressure were taken from the Marketing Scales Handbook (Bruner et al. 2001) and partly from a modified version of a labeling study conducted by Buxel and Schulz (2010). However, most of the items had to be developed by the authors due to the explorative character of the study.

Initially, expert pre-tests were carried out in order to carefully examine whether the items loaded were the intended latent variables. Ten colleagues involved in similar research fields were asked to allocate the items, which were mixed in order of their latent variables. Those items not allocated to the same latent variable by the majority of experts were removed. Finally, before the field phase, pre-tests were carried out with consumers of differing age, sex, education and net household income.

The field phase took place in September and October 2014. All statements were presented randomly to the participants to prevent sequence effects.

The data was analyzed using SmartPLS statistical software, Version 3.1 and by applying structural equation modeling methods (Ringle et al. 2015). The analysis of PLS models contains two steps which are an assessment of the reliability and validity of the measurement model (outer model) and an assessment of the structural model's fit (inner model) (Hair et al. 2011).

[1] For the choice sets and the corresponding detailed information given to the respondents see Figures A1 and A2 of the Appendix.

The moderator modeling framework described above was then built into SmartPLS. Afterwards, the PLS Algorithm, bootstrapping, blindfolding procedures and the multi group analysis (MGA; Sarstedt et al. 2011) were conducted.

Results

Sociodemographic Characteristics of the Sample

Due to the set quotas, the sample is in accordance with the distribution of age, sex, regional distribution, education and net household income in Germany. Furthermore, this is also valid for both sub-samples as the quotas were set within each of them. The average age is forty-four years, and 49.2% of the participants are male. The regional distribution corresponds to the German population. Of the respondents, 30.1% have a net household income of less than € 1,500 per month. Only the education levels were not perfectly met, as the higher education level is slightly overrepresented. However, this is not unusual for online surveys as Granello and Wheaton (2004) have shown, and the advantages of online surveys outweigh this single disadvantage. The complete results including the results for the distribution in Germany can be found in Table A1 of the Appendix.

Satisfaction with an Animal Welfare Label

To capture first impressions of satisfaction with the respective label, statements were evaluated descriptively and mean values calculated for each sub-sample. In order to analyze whether there are significant differences between the binary and the multi-level labels sub-sample t-tests were performed. The results are shown in Table 1 below.

It can be seen that for each statement the mean value is higher for the multi-level split. With the exception of three items (1: “I think the animal welfare label is reliable”; 7: “Products with this label would attract my attention in the supermarket”; 10: “I would look for such products at the weekly market”) the differences are statistically significant. In more detail, consumers in the second split (=multi-level-label) have a better idea of what the label is based on than in the first split (2; μ Split 1 = 0.46; μ Split 2 = 0.58). The same applies to Item 3; that the animal welfare label facilitates the comparison between products (μ Split 1 = 0.41; μ Split 2 = 0.54). Participants in the multi-level split are also of the opinion that the label provides precise information without being forced to read it for a long time (4; μ Split 1 = 0.34; μ Split 2 = 0.53) and they also think that the label gives the opportunity to learn more about a product without taking up too much time (5; μ Split 1 = 0.39; μ Split 2 = 0.57). In addition consumers in the multi-level split are more interested in products with this type of label than in the binary split (6; μ Split 1 = 0.55; μ Split 2 = 0.67). Furthermore and as a matter of choice participants in Split 2 would look for products with such a label in the supermarket or at the butcher (8; μ Split 1 = 0.35; μ Split 2 = 0.47; 9; μ Split 1 = 0.26; μ Split 2 = 0.36).

Table 1. Satisfaction with an animal welfare label

	Statement	I totally disagree (-2)	I disagree (-1)	Partly/ partly (0)	I agree (1)	I totally agree (2)	MV (SD)
1	I think the animal welfare label is reliable.	3.6	9.0	45.2	33.8	8.4	0.34 (0.888)
		3.1	9.1	41.1	37.2	9.4	0.41 (0.895)
2	I can imagine what the animal welfare label is based on.***	3.2	11.3	32.1	43.2	10.2	0.46 (0.933)
		2.7	9.1	28.4	46.5	13.3	0.58 (0.925)
3	The animal welfare label facilitates the comparison between products.***	4.0	13.9	31.0	39.8	11.4	0.41 (0.992)
		3.9	9.4	30.2	41.6	14.9	0.54 (0.985)
4	The animal welfare label provides me precise information without being forced to read for a long time.***	4.2	13.2	36.3	36.9	9.5	0.34 (0.965)
		3.3	9.4	31.6	42.3	13.4	0.53 (0.950)
5	The animal welfare label gives me the opportunity to learn more about a product without requiring me to read for a long time.***	3.4	13.4	34.6	38.0	10.6	0.39 (0.962)
		3.4	8.8	29.4	43.8	14.6	0.57 (0.957)
6	I am interested in products with such a label.**	3.7	10.7	28.9	40.0	16.6	0.55 (1.009)
		4.6	6.7	26.0	43.1	19.7	0.67 (1.012)
7	Products with this label would attract my attention in the supermarket.	3.8	14.2	32.5	38.9	10.6	0.38 (0.982)
		3.9	10.9	33.9	39.2	21.1	0.45 (0.971)
8	I would look for products with such a label in the supermarket.**	5.6	15.0	30.5	37.0	11.9	0.35 (1.050)
		4.8	12.0	30.9	35.7	16.5	0.47 (1.054)
9	I would look for such products at the butcher.**	6.0	17.3	32.6	33.5	10.7	0.26 (1.054)
		6.0	14.0	33.0	31.7	15.3	0.36 (1.085)
10	I would look for such products at the weekly market.	7.5	18.8	32.6	31.2	10.0	0.17 (1.082)
		7.6	15.3	34.4	29.5	13.2	0.25 (1.103)

Notes. The first line indicates sub-sample one with the binary label; the second line indicates sub-sample two (multi-level)
* = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level
Source. Authors' own calculations

Evaluation of the Measurement Model

In the following sections, the results of the PLS analysis are presented. Before starting the analyses, we conducted a confirmatory factor analysis in order to identify all the relevant items for a latent variable (cf. Appendix, Table C). All items that had an outer loading of less than 0.4

were removed from the latent variables (4 items[2]) (ibid.). Apart from a few exceptions (one loading each on *satisfaction*, *involvement*, *comprehensibility*, *time pressure* and two loadings each on *trust* and *use*; all are > 0.4 but < 0.7), all the factor loadings on each latent variable are above a threshold of 0.7 or higher (Hair et al. 2013). Removing these items did not improve internal consistency reliability, and so they were included in the analyses (ibid.).

For internal consistency reliability, we applied both Cronbach's Alpha and composite reliability. This was done as Cronbach's Alpha usually tends to underestimate internal consistency reliability (Hair et al. 2013), whilst composite reliability is possibly a better indicator as it takes into account outer loadings of the indicator variables (Hensler et al. 2009, Hair et al. 2011). The recommended threshold of 0.7 or above was reached for both reliability coefficients (Nunnally and Bernstein 1994). The convergent validity was measured by the average variance extracted (AVE), which is comparable to the proportion of explained variance in the factor analysis and should be higher than 0.5 so that it explains more than half of the indicators' variance on average (Fornell and Larcker 1981). This criterion is fulfilled by the model. Table 2 contains the results for Cronbach's Alpha, the composite reliability and the average variance extracted in detail.

Table 2. Evaluation of the measurement model

Variables	Number of items	CRA (>= 0.7)	CR (>=0.7)	AVE (>=0.5)
Satisfaction	10	0.926	0.938	0.602
Comprehensibility	6	0.830	0.863	0.514
Involvement	8	0.873	0.900	0.533
Time pressure	6	0.857	0.866	0.522
Trust	7	0.870	0.900	0.569
Use	5	0.806	0.868	0.572

Notes. CRA = Cronbach's Alpha, CR = Composite reliability, AVE = Average variance extracted
Source. Authors' own calculations

For the assessment of the discriminant validity, the Fornell-Larcker criterion, and the cross loading should be considered. The Fornell-Larcker criterion requires that a latent variable should explain the variance of its own indicators better than the variance of other latent variables (Fornell and Larcker 1981). This criterion is met (see Appendix, Table A3). For the cross-loading criterion, an indicator's loading on its assigned latent variable should be higher than its loading on all other variables (ibid.). The data analysis revealed that there was no evidence of any cross-loadings (data not provided but available on request).

Evaluation of the Structural Model

A number of quality criteria were applied for the evaluation of the structural model. Firstly, the coefficient of determination, R-squared, was found to be 0.499 for the endogenous variable

[2] *Labels on packages often present food better than it is in reality. (recoded)*; I miss important information on labels on food packages.; *I find details on food package well explained. (recoded)*; For many labels, I cannot understand the meaning.

satisfaction. This means that the five latent variables moderately explain 49.9% of the variance in *satisfaction* (Hair et al. 2011). Secondly, the predictive relevance, f-squared, was determined, whereby we found a moderate effect for *trust* (0.160). All the other four latent variables showed weak effects (< 0.15) (Cohen 1988). Thirdly, we evaluated the size and significance of all the path coefficients in the complete model and in the two sub-samples (binary vs. multi-level label) (see below).

The bootstrap analysis showed that all five latent variables – *comprehensibility* ($p \leq 0.10$), *involvement* ($p \leq 0.1$), *time pressure* ($p \leq 0.05$), *trust* ($p \leq 0.1$) and *use* ($p \leq 0.1$) – have a significant effect on *satisfaction* in the case of the complete sample. These results indicate that the influence of *trust* is the strongest (0.389) on consumer *satisfaction* with the perspective labeling system, followed by *use* (0.266), *involvement* (0.141) and *comprehensibility* (0.043). *Time pressure* was found to have a negative influence (-0.074) on *satisfaction*.

Multi-Group Analysis

A comparison of the binary and the multi-level label split revealed slight differences. R-squared for the binary split is 0.513, for the multi-level split 0.497. The composite reliability criterion was met for all five variables in both splits. However, the AVE for *comprehensibility* in the binary split is slightly smaller than the threshold of 0.5 (0.491). For all the other latent variables, the AVE is larger than 0.5. All the results for the variables for both splits can be seen in detail in Table A4 in the Appendix.

The multi-group analysis (MGA) also showed differences in the path coefficients. For the binary split, two of the five latent variables were found not to have a significant effect (*time pressure* and *comprehensibility*). In contrast, all five latent variables showed significant results for the multi-level split.

To summarize, the hypotheses H2, H4, and H5 can be confirmed: *involvement*, *trust*, and *use* have a significant influence on consumers' *satisfaction* with labels for both splits. For *time pressure* (H3) and *comprehensibility* (H1), the hypotheses could only be verified for the multi-level split. An overview of the results is provided in Figure 2.

The bootstrap for PLS-MGA (Henseler 2007) revealed that two of the five latent variables – namely *trust* and *time pressure* – have significant differences between the consumers who saw the binary label and those who saw the multi-level label ($p \leq 0.1$). *Comprehensibility*, *involvement*, and *use* do not show any significant differences whether the respondents were confronted with a binary or a multi-level label. Consequently, the significant differences confirm hypotheses H3 and H4. The detailed findings are presented in Table 3.

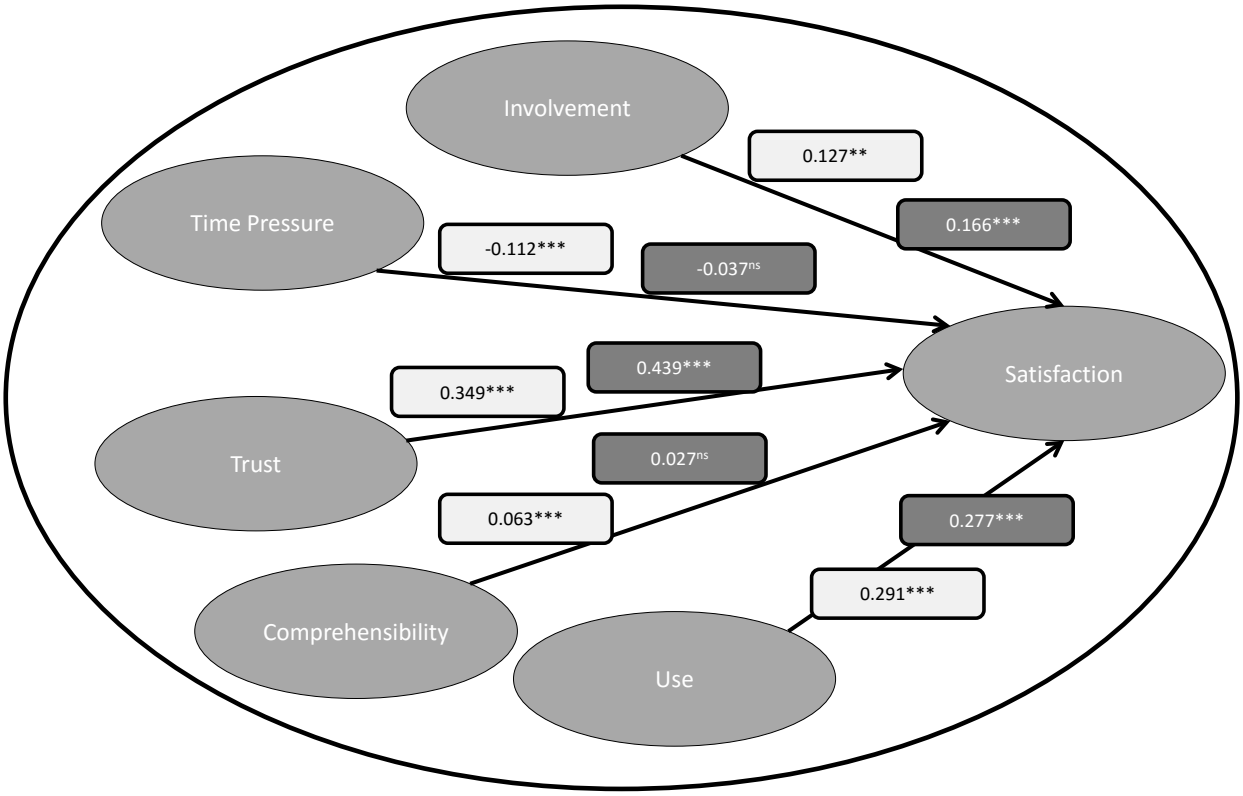

Figure 2. Results of the path coefficients

Notes: * = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level; ns = not significant; dark grey boxes = binary split; light grey boxes = multi-level split; for detailed results see Tables A5 and A6 of the Appendix

Source. Authors' own calculations and illustration

Table 3. Multi-group comparison results

Relationship	Comparison	|diff|	PLS-MGA (P-Henseler)
comprehensibility → satisfaction	binary vs. multi-level label	0.036	0.799
involvement → satisfaction	binary vs. multi-level label	0.039	0.253
time pressure → satisfaction	binary vs. multi-level label	0.075	0.069
trust → satisfaction	binary vs. multi-level label	0.091	0.080
use → satisfaction	binary vs. multi-level label	0.064	0.819

Source. Authors' own calculations

Discussion and Conclusion

The results of the present study on the five latent variables and their effect on consumer satisfaction with respect to the two different labeling schemes (binary and multi-level) provided interesting insights. At first sight, significant differences for seven out of ten statements building the latent variable *satisfaction* indicate that consumers might be more satisfied with a multi-level label than with a binary label.

There were no statistical differences between the two labeling schemes for the latent variables *comprehensibility*, *involvement*, and *use*. In contrast, there were differences for both *time pressure* and *trust*. The group comparison indicated that trust as a precondition is more necessary for a binary label whereas high time pressure reduces satisfaction with a multi-level label.

There could be several explanations for the lack of statistical difference between the two sub-samples for *comprehensibility*. First, it might be the case that the multi-level label was no more easily understood than the binary-level label. However, the influence of *comprehensibility* on *satisfaction* was not significant for the binary split and when only the multi-layer split is considered on its own then *comprehensibility* significantly increases *satisfaction*. This is an indication that the multi-level labeling schemes might indeed be superior in terms of *comprehensibility*. Secondly, the non-significant result in the MGA might be explained by the information overload theory (Kolodinsky 2012; Kroeber-Riel and Esch 2004; van Kleef et al. 2008; Verbeke 2005), as a five-level label such as the one used in this study is rather complex. Accordingly, this study's multi-level label may have provided too much information so that the probands applied heuristics rather than trying to understand the underlying principles of the labeling scheme (Schulte-Mecklenbeck et al. 2013). However, to verify this assumption, further research needs to be carried out.

If five levels are too many, the vital factor of how to design the labeling scheme would not be how much information is included but whether or not consumers are confronted with too much information. Therefore, *comprehensibility* remains a crucial factor for food labeling in general (Aarset et al. 2004; Grunert and Wills 2007). Food packages are often cluttered and not user-friendly. Producers should focus instead on those essential features they want to emphasize in the package design. However, one argument against the idea that a five-level label is too complex to have high *comprehensibility* is that the worldwide five-level labeling system works for hotel classification. Further studies analyzing *comprehensibility* should be aimed at detecting the optimal number of grades for a multi-level food label.

There was also no statistical difference between the splits for the second latent variable *involvement* although *involvement* was seen to have an influence on consumer *satisfaction* when each split was considered separately. It can, therefore, be concluded that *involvement* has an influence, but its strength is similar in both of the two labeling schemes. This may be explained by considering the level of involvement as an attitude or part of an individual's character rather than an influence on a labeling scheme. So whether consumers make the effort to gather further information on a labeling scheme depends on their involvement characteristics. This explanation for the current result would be in line with the outcomes of Spiller (2010) and Espejel et al. (2009) who also stated that consumers either have high involvement or not in food products.

The fact that the PSL-MGA for *use* is also not significant is unsurprising considering the result for *comprehensibility*. *Use* is closely connected with the label being understandable or not. Labels are generally considered to be important for consumers (Besler et al. 2012, Buxel and Schulz 2010). However, it is also entirely possible that a multi-level labeling scheme does not outclass a binary label in terms of impact on the user. Nevertheless, the literature still shows that label design remains a crucial factor in its use (Besler et al. 2012, Silayoi and Speece 2004).

A significant difference between the two food labeling systems can be found for *time pressure*. *Time pressure* was found to lower satisfaction with a multi-level label compared to a binary label as the value for the multi-level split is more negative (-0.112) than the result for the binary split (-0.037). This could be explained by the fact that a multi-level labeling scheme is presently unknown in the food sector. The pro-bands (i.e. the consumers) would have to become accustomed to such a labeling system, especially as weekly food shopping is characterized by habit purchases (Silayoi and Speece 2004, Schulte-Mecklenbeck et al. 2013). Furthermore, this might be supported by the argument that the multi-level seems more complex for consumers at first sight. However, once a multi-level labeling scheme becomes more familiar to consumers, *time pressure* might have an opposite effect. That is if a product's details are shown transparently on food packaging so consumers would not have to search for additional information on different binary labels (e.g. about the governmental organic label and an organic label allocated by an organic farming association). A multi-level label can indicate differences in such labels without the requirement of additional information. These assumptions should be verified by further qualitative research approaches. However, this result does emphasize the need to take great care in introducing a multi-level labeling system. No matter whether the label is allocated by the state or by a manufacturer, its introduction should be accompanied by marketing campaigns such as advertising on nationwide TV to make both the labeling system and the label itself popular.

Although a significant difference was found for *trust* between *satisfaction* with the multi-level (0.349) and the binary label (0.439), *trust* still seems to have less influence on *satisfaction* with the multi-level label than the binary label. This can be explained by the fact that a multi-level label displays more detailed information than a binary label and might thus inspire more confidence. As *trust* is an essential component for a label's success (Fenko et al. 2016, Herrera and Blanco 2011, Nocella et al. 2010, 2014), it is a good indicator for a multi-level label's potential market impact as the multi-level label seems to be more trustworthy. Nevertheless, it is important to investigate the underlying reasons for this difference in future qualitative research. *Trust* might also be influenced by the label allocator or the label's design. It would be most important to learn about these influences before the setting up and launching of a multi-level food label. This result is especially interesting for producers. In times when consumers have little trust in the food sector, a multi-level label seems to provide an excellent opportunity to enhance reputation.

As ever in empirical research, this study also has its limitations. It was the first one of its kind, and its design could be improved by experiences gained during the explorative survey into further studies. Another challenge for the preparation of a multi-level label launch is the pricing of the different levels. For this purpose, progressive as well as retrograde cost analyses should be calculated and matched with willingness-to-pay studies for ethical products. In addition, research should be carried out for different ethical products in order to gain deeper consumer insights into the grading of ethically labeled products.

The results indicate that a multi-level label has to be introduced carefully, and its design needs to be thought through in great detail. Grading should also be consistent with a multi-level labeling system, e.g. only stars, but not "premium stars", "star plus" or "star superior" in order to prevent consumer confusion. However, such a label can also be used to improve ethical food products' labeling as differentiated product characteristics are displayed to consumers more transparently.

Multi-levelling, therefore, provides producers with a good tool to highlight the advantages of a premium quality product. Are there also options for more price sensitive consumers who do not want to pay a high surplus for ethical attributes? If producers do not agree to consistent labeling schemes for ethical products or if there is failure of agreement to fund a marketing campaign, it might be advisable that the state is the allocator. The success of any multi-level labeling scheme will depend on transparency and communication.

Acknowledgement

This study was supported by the German Academic Exchange Service in the form of a short-term doctoral research scholarship.

References

Aarset, B., S. Beckmann, E. Bigne, M. Beveridge, T. Bjorndal, J. Bunting, P. McDonagh, C. Mariojouls, J. Muir, A. Prothero, L. Reisch, A. Smith, R. Tveteras and J. Young. 2004. The European consumers' understanding and perceptions of the "organic" food regime: The case of aquaculture. *British Food Journal* 106(2):93–105.

Abrate, G., A. Capriello and G. Fraquelli. 2011. When quality signals talk: Evidence from the Turin hotel industry. *Tourism Management* 32(4):912–921.

Aertsens, J., K. Mondelaers, and G. van Huylenbroeck. 2009. Differences in retail strategies on the emerging organic market. *British Food Journal* 111(2):138–154.

Akerlof, G.A. 1970. The market for "lemons": Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics* 84(3):488–500.

Alberini, A., M. Bareit and M. Filippini. 2016. What is the effect of fuel efficiency information on car prices? Evidence from Switzerland. *The Quarterly Journal of the IAEE's Energy Economics Education Foundation* 37(3):315–342.

Andrews, J.C., S. Burton and J. Kees. 2011. Is simpler always better? Consumer evaluations of front-of-package nutrition symbols. *Journal of Public Policy & Marketing* 30(2):175–190.

Besler, H.T., Z. Buyuktuncer, and M.F. Uyar. 2012. Consumer understanding and use of food and nutrition labeling in Turkey. *Journal of Nutrition Education and Behavior* 44(6):584–591.

Bokkers, E.A.M., I.J.M. de Boer, and P. Koene. 2011. Space needs of broilers. *Animal Welfare* 20 (4):623–632.

Bruner, G.C., K.E. James, and P.J. Hensel. 2001. Marketing Scales Handbook: A Compilation of Multi-Item Measures. Volume III. *American Marketing Association*. USA, Chicago.

Buxel, H., and S. Schulz. 2010. Akzeptanz und Nutzung von Güte- und Qualitätssiegeln auf Lebensmitteln [Acceptance and usage of quality labels on food]. Fachhochschule Münster. https://www.fh-muenster.de/fb8/downloads/buxel/10_Studie_Lebensmittelsiegel.pdf. [accessed February 2, 2015].

Caputo, V., A. Vassilopoulos, R.M. Nayga, and M. Canavari. 2013. Welfare effects of food miles labels. *Journal of Consumer Affairs* 47(2):311–327.

Caswell, J.A., and D.I. Padberg. 1992. Toward a more comprehensive theory of food labels. *American Journal of Agricultural Economics* 74(2):460–468.

Caswell, J.A., and E.M Mojduszka. 1996. Using informational labeling to influence the market for quality and in food products. *American Journal of Agricultural Economics* 78(5):1248–1253.

Clark, B., Stewart, G.B., Panzone, L.A., Kyriazakis, I. and L.J. Frewer. 2016. A systematic review of public attitudes, perceptions and behaviours towards production diseases associated with farm animal welfare. *Journal of Agricultural and Environmental Ethics* 29(3):455–478.

Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Darby, M.R., and E. Karni. 1973. Free competition and the optimal amount of fraud. *Journal of Law and Economics* 16(1):67–88.

Darian, J.C., and J. Cohen. 1995. Segmenting by consumer time shortage. *Journal of Consumer Marketing* 12: 32–44.

De Jonge, J., and H.C.M. van Trijp. 2013. The impact of broiler production system practices on consumer perceptions of animal welfare. *Poultry Science* 92(12):3080–3095.

Eichholtz, P., N. Kok, and J.M. Quigley. 2010. Doing well by doing good? Green office buildings. *American Economic Review* 100(5):2494–2511.

Espejel, J., C. Fandos and C. Flavián. 2009. The influence of consumer involvement on quality signals perception: An empirical investigation in the food sector. *British Food Journal* 111 (11):1212-1236.

Espinet, J. M., M. Saez, G. Coenders, and M. Fluvià. 2003. Effect on prices of the attributes of holiday hotels: A hedonic prices approach. *Tourism Economics* 9(2):165–177.

Fenko, A., L. Kersten and S. Bialkova. 2016. Overcoming consumer scepticism toward food labels: The role of multisensory experience. *Food Quality and Preference* 48:81–92.

Feunekes, G.I.J., I.A. Gortemaker, A.A. Willems, R. Lion, and M. van den Kommer. 2008. Front-of-pack nutrition labeling: Testing effectiveness of different nutrition labeling formats front-of-pack in four European countries. *Appetite* 50(1):57–70.

Fisher, C., T.P. Lyon, and A. Arbor. 2013. A theory of multi-tier ecolabels. Paper presented at the 134th annual meeting of European Association of Agricultural Economists (EAAE) Paris, France. March 22, 2013.

Fornell, C., and D.F. Larcker. 1981. Evaluation structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18(1):39–50.

Galarraga, I., González-Eguino, M. and A. Markandya. 2011. Willingness to pay and price elasticities of demand for energy-efficient appliances: Combining the hedonic approach and demand systems. *Energy Economics* 33(1):66–74.

Galarraga, I., A. Ramos, J. Lucas and X. Labandeira. 2014. The price of energy efficiency in the Spanish car market. *Transport Policy* 36:272–282.

Granello, D.H., and J.E. Wheaton. 2004. Online data collection: strategies for research. *Journal Counseling & Development* 82(4):387–393.

Grebitus, C., Steiner, B. and M. Veeman. 2015. The roles of human values and generalized trust on stated preferences when food is labeled with environmental footprints: Insights from Germany. *Food Policy* 52: 84–91. doi: 10.1016/j.foodpol.2014.06.011.

Grunert, K.G., 2002. Current issues in the understanding of consumer food choice. *Trends in Food Science & Technology* 13(8):275–285.

Grunert, K. G., and J.M. Wills. 2007. A review of European research on consumer response to nutrition information on food labels. *Journal of Public Health* 15(5):385–399.

Hair, J.F., G.T.M. Hult, C.M. Ringle, and M. Sarstedt. 2013. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM).* Thousand Oaks: Sage.

Hair, J.F., Ringle, C.M. and M. Sarstedt. 2011. PLS-SEM Indeed a Silver Bullet. *Journal of Marketing Theory and Practice* 19(2):139–151.

Henseler, J., 2007. *A new and simple approach to multi-group analysis in partial least squares path modeling*, paper presented at the 5th international symposium on PLS and related methods (PLS'07), edited by H. Martens and T. Naes, 104–107. Oslo, Norway. September.

Hensler, J., Ringle, C.M. and R.R Sinkovics. 2009. The use of partial least squares path modeling in international marketing. *Advances in International Marketing* 20:277–319.

Herrera, C.F., and C.F. Blanco. 2011. Consequences of consumer trust in PDO food products: the role of familiarity. *Journal of Product & Brand Management* 20(4):282–296.

Hoek, J., N. Roling and D. Holdsworth. 2013. Ethical claims and labeling: An analysis of consumers' beliefs and choice behaviors. *Journal of Marketing Management* 29(7-8):772–792.

IGD.com. 2010. Ethical and sustainable shopping. http://shoppervista.igd.com/Hub.aspx?id=32&tid=4&m=1&rptid=139.

Jahn, G., Schramm, M. and A. Spiller. 2005. The reliability of certification. *Journal of Consumer Policy* 28 (1):53–73.

Kelly, B., Hughes, C., Chapman, K., Louie, J.C.-Y., Dixon, H., Crawford, J., King, L., Daube, M. and T. Slevin. 2009. Consumer testing of the acceptability and effectiveness of front-of-pack food labeling systems for the Australian grocery market. *Health Promotion International* 24 (2):120–129.

Kolodinsky, J. 2012. Persistence of Health Labeling Information Asymmetry in the United States: Historical Perspectives and Twenty-First Century Realities. *Journal of Macromarketing* 32 (2):193–207.

Kroeber-Riel, W., and F.-R. Esch. 2004. *Strategie und Technik der Werbung: Verhaltenswissenschaftliche Ansätze [Strategy and technology in advertising: Behavioural approches]*. Kohlhammer GmbH, Stuttgart.

Krystallis, A., and G. Chryssohoidis. 2005. Consumers' willingness to pay for organic food. *British Food Journal* 107 (5):320-343.

Lang, T., and M. Heasman. 2004. *Food wars. The global battle for mouths, minds and markets.* Earthscan Publications, London.

Nelson, P. 1970. Information and consumer behaviour. *Journal of Political Economy* 78(2):311–329.

Nocella, G., Hubbard, L. and R. Scarpa. 2010. Farm animal welfare, consumer willingness to pay, and trust: Results of a cross-national survey. *Applied Economic Perspectives and Policy* 32 (2):275–297.

Nocella, G., Romano, D. and G. Stefani. 2014. Consumers' attitudes, trust and willingness to pay for food information. *International Journal of Consumer Studies* 38:153-165.

Nunnally, J.C., and I.H. Bernstein. 1994. *Psychometric Theory.* 3rd edition. MacGraw-Hill, New York.

Pouta, E., J. Heikkila, S. Forsman-Hugg, M. Isoniemi, and J. Makela. 2010. Consumer choice of broiler meat: The effects of country of origin and production methods. *Food Quality and Preference* 21(5):539–546.

Ringle, C.M., Wende, S. and J.-M. Becker. 2015. SmartPLS 3. Bönningstedt: *SmartPLS.* http://www.smartpls.com.

Roberto, C., M. Bragg, M.B. Schwartz, M.J. Seamans, A. Musicus, N. Novak, and K.D. Brownell. 2012. Facts up front versus traffic light food labels: A randomized controlled trial. *American Journal of Preventive Medicine* 43(2):134–141.

Roosen, J., A. Bieberstein, S. Blanchemanche, E. Goddard, S. Marette, and F. Vandermoere. 2015. Trust and willingness to pay for nanotechnology food. *Food Policy* 52:75-83.

Sarstedt, M., J. Henseler, and C.M. Ringle. 2011. Multi-group analysis in partial least squares (PLS) path modeling: alternative methods and empirical results. *Advances in International Marketing* 22:195-218.

Schröder, M.J.A., and M. G. McEachern. 2004. Consumer value conflicts surrounding ethical food purchase decisions: a focus on animal welfare. *International Journal of Consumer Studies* 28 (2):168–177.

Schulte-Mecklenbeck, M., M. Sohn, E. de Bellis, N. Martin,. and R. Hertwig. 2013. A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite* 71:242–251. doi: 10.1016/j.appet.2013.08.008.

Scott, V., and A. F. Worseley. 1994. Ticks, claims, tables and food groups: a comparison of nutrition labeling. *Health Promotion International* 9(1):27–37.

Silayoi, P., and M. Speece. 2004. Packaging and purchase decisions. *British Food Journal* 106 (8):607–628.

Spiller, A. 2010. *Marketing Basics. Ein Online-Lehrbuch [Marketing basics. An online textbook]*. Georg-August University, Göttingen.

Su, C., and L. Sun. 2007. Taiwan's hotel rating system: A service quality perspective. *Cornell Hotel and Restaurant Administration Quarterly* 48(4):392–401.

Talebi, A., M.A.G. von Keyserlingk, E. Telezhenko, and D.M. Weary. 2014. Reduced stocking density mitigates the negative effects of regrouping in dairy cattle. *Journal of Dairy Science* 97(3):1358–1363.

Turnbull, J.F., A. Bell, C.E. Adams, J. Bron, and F. Huntingford. 2005. Stocking density and welfare of cage-farmed Atlantic salmon: application of a multivariate analysis. *Aquaculture* 243(1-4):121–132.

United States Department of Agriculture, Agricultural Marketing Service. (USDA–AMS). 2012. Labeling Organic Products. http://www.ams.usda.gov/AMSv1.0/getfile?dDocName=STELDEV3004446.

van der Merwe, D., E. Kempen, S. Breedt, and H. de Beer. 2010. Food choice: student consumers' decision-making process regarding food products with limited label information. *International Journal of Consumer Studies* 34(1):11–18.

Van Kleef, E., H. van Trijp, F. Paeps, and L. Fernández-Celemín. 2008. Consumer preferences for front-of-pack calories labeling. *Public Health Nutrition* 11(2):203–213.

Vanhonacker, F., W. Verbeke, E. Van Poucke, S. Buijs. and F.A.M. Tuyttens. 2009. Societal concern related to stocking density, pen size and group size in farm animal production. *Livestock Science* 123 (1):16–22.

Vecerek, V., S. Grbalova, E. Voslarova, B. Janackova, and M. Malena. 2006. Effects of travel distance and the season of the year on death rates of broilers transported to poultry processing plants. *Poultry Science* 85(11):1881–1884.

Verbeke, W., 2005. Agriculture and the food industry in the information age. *European Review of Agricultural Economics* 32 (3):347–368.

Verbeke, W., Z. Pieniak, L. Guerrero, and M. Hersleth. 2012. Consumers' awareness and attitudinal determinants of European Union quality label use on traditional foods. *Bio-based and Applied Economics* 1(2):213–229.

Warde, A., 1999. Convenience food: space and timing. *British Food Journal* 101 (7):518-527.

Weinrich, R., and A. Spiller. 2016. Developing food labeling strategies: Multi-level labeling. *Journal of Cleaner Production* 137:1138–1148.

Weinrich, R., A. Franz, and A. Spiller. Forthcoming. Multi-level labeling: Too complex for consumers? *Economica Agro-Alimentare*.

Williams, S., and K.W. Mummery. 2012. Characteristics of consumers using 'better for you' front-of-pack food labeling schemes - an example from the Australian Heart Foundation Tick. *Public Health Nutrition* 16 (12):2265-2272.

Appendix

It is about animal welfare in livestock farming. The label is controlled and granted by the Federal Republic of Germany. The requirements go beyond the legal minimum standards of livestock farming. Here are two examples:

In the traditional, conventional pig fattening there are **0.75 sqm of space** available per pig by law. The transport time to the slaughterhouse may take **8 hours**.

A fattening pig standing in a stall that is certified with the label has are **1.50 sqm of space** available. The transportation duration to the slaughterhouse may not take more than **4 hours.**

Figure A1. Choice set and label information for the binary label
Source. Weinrich and Spiller.

It is about animal welfare in livestock farming. The label is controlled and granted by the Federal Republic of Germany. The requirements go beyond the legal minimum standards of livestock farming. Here are two examples:

5 stars: **2.00 sqm** space and maximum **2 hours** transport time to slaughterhouse
4 stars: **1.75 sqm** space and maximum **3 hours** transport time to slaughterhouse
3 stars: **1.50 sqm** space and maximum **4 hours** transport time to slaughterhouse
2 stars: **1.25 sqm** space and maximum **5 hours** transport time to slaughterhouse
1 star: **1.00 sqm** space and maximum **6 hours** transport time to slaughterhouse
without label: **0.75 sqm** Platz and maximum **8 hours** transport time to slaughterhouse

Figure A2. Choice set and label information for the multi-level label
Source. Weinrich and Spiller

Table A1. Sample characterization

Variable	Description	Frequency (%) Sub-sample 1	Frequency (%) Sub-sample 2	Frequency (%) Germany
Age	16 to 29	22.1	20.7	22.3
	30 to 39	17.0	17.3	17.3
	40 to 49	21.1	20.7	21.1
	50 to 59	23.1	23.7	22.8
	60 to 69	16.6	17.7	16.6
Gender	Male	49.5	49.0	48.8
	Female	50.5	51.0	51.2
Region	North	15.7	16.5	16.1
	South	28.0	25.2	28.7
	East	21.7	21.2	21.0
	West	34.6	37.1	34.2
Education level	No qualification	1.2	0.5	4.8
	Primary school	21.1	19.0	30.7
	Secondary school	36.9	37.6	32.3
	Technical college qualification	9.8	9.9	7.7
	German equivalent to A-levels	31.1	33.0	24.5
Net household income(€)	Less than 500	3.4	2.9	2.1
	500-899	8.1	7.7	10.4
	900-1,499	19.2	19.1	21.8
	1,500-1,999	19.0	19.7	16.4
	2,000-2,599	17.9	19.0	15.6
	2,600-3,199	12.6	13.4	10.9
	More than 3,200	19.8	18.2	22.8

Source. Authors' calculations on the basis of preliminary results of the German census 2011, census data in the version of 10/04/2014 (Federal Statistical Office 2014)

Table A2. Measurement items for the variables in the research model

Variables and Measurement Items[1]	Mean	Standard Deviation	Factor loading[2]
Satisfaction			
I believe the animal welfare label is reliable.	0.37	0.892	0.738
The animal welfare label makes it easier to compare products.	0.47	0.988	0.823
I am interested in products with such a label.	0.61	1.007	0.806
I can imagine what stands behind the animal welfare label.	0.52	0.932	0.657
The animal welfare label provides me with exact information without me having to read too much.	0.44	0.959	0.762
The animal welfare label gives me the opportunity to know more about a product without me having to inform myself more comprehensively.	0.48	0.960	0.760
Products with this label would attract my attention in the supermarket.	0.42	0.975	0.744
I would look for products with such a label in the supermarket.	0.41	1.051	0.846
I would look for products with such a label at the butchers.	0.31	1.069	0.810
I would look for products with such a label at the weekly market.	0.21	1.093	0.799
Comprehensibility			
The many different labels on food packages confuse me.	0.13	0.935	0.667
I do not get the meaning of many labels.	0.61	0.782	0.785
Information about food confuses me.	0.03	0.876	0.704
I do not know what lies behind many of the labels.	0.71	0.803	0.786
Information on food packages is often incomprehensible.	0.51	0.846	0.723
Information about food is too complicated.	0.18	0.888	0.621
Involvement			
It is important for me to get good quality food.	0.97	0.721	0.752
When it comes to buying products, I try to make a perfect choice.	0.65	0.811	0.742

Table A2. *–Continued*

Variables and Measurement Items[1]	Mean	Standard Deviation	Factor Loading[2]
I usually try to buy the very best quality.	0.49	0.868	0.791
I always try to choose the very best quality products.	0.49	0.839	0.786
I am interested in food.	0.71	0.849	0.743
I would like to know which ingredients are contained in food.	0.89	0.894	0.731
Concerning food I just want to fill my stomach.	0.60	1.007	0.562
I think a lot about food.	0.15	0.992	0.708
Time Pressure			
I find myself pressed for time when I go grocery shopping.	-0.51	0.947	0.672
I am in a hurry when I do my grocery shopping.	-0.43	0.951	0.681
I have only a limited amount of time to finish my grocery shopping.	-0.22	1.076	0.622
I quickly finish my grocery shopping because I have other things to do.	-0.09	0.997	0.749
In everyday life, I take the time to busy myself with the information on the food package when doing the shopping.	-0.18	0.948	0.772
I have more than enough time to complete my weekly grocery shopping.	-0.48	0.955	0.819
Trust			
I trust the governmental food control.	0.09	0.929	0.602
I trust food control which is carried out by private companies.	-0.36	0.930	0.529
Information on food packages is an important purchase aid for me because I trust it.	0.17	0.904	0.776
I have considerable confidence in the labels on food packages.	-0.13	0.882	0.856
I have more trust in food with labels than in alternative products without labels.	0.009	0.940	0.806
Labels help me to recognize the quality of food.	0.24	0.886	0.848
Labels are a good thing.	0.43	0.824	0.800

Table A2. *–Continued*

Variables and Measurement Items[1]	Mean	Standard Deviation	Factor Loading[2]
Use			
I particularly notice food with labels when shopping.	0.08	0.923	0.793
I deliberately choose products with labels.	-0.16	0.882	0.874
If a food product has a label and another not, I will choose the product with the label.	-0.10	0.851	0.820
Yet another label – nobody needs this.	0.01	1.004	0.614
I know a lot about labels on food packages.	-0.57	0.833	0.646

Notes. [1]Respondents answered on a five-point Likert scale; [2]Results of the PLS confirmatory factor analysis, italicized items were recoded.
Source. Authors' own calculations

Table A3. Discriminant validity analysis based on the Fornell-Larcker criterion

Variables	Satisfaction	Comprehensibility	Involvement	Time Pressure	Trust	Use
Satisfaction	**0.776**					
Comprehensibility	-0.136	**0.726**				
Involvement	0.458	-0.076	**0.730**			
Time pressure	-0.345	0.140	-0.462	**0.734**		
Trust	0.640	-0.249	0.373	-0.294	**0.755**	
Use	0.625	-0.234	0.533	-0.362	0.696	**0.757**

Notes. Diagonal values in bold are the square root of the average variance extracted.
Source. Authors' calculations

Table A4. Label-specific results

		Binary Label	Multi-Level Label
Latent Variables			
Comprehensibility	CR	0.851	0.877
	AVE	0.491	0.544
Involvement	CR	0.987	0.903
	AVE	0.525	0.541
Time Pressure	CR	0.897	0.903
	AVE	0.530	0.545
Trust	CR	0.871	0.877
	AVE	0.562	0.578
Use	CR	0.865	0.904
	AVE	0.567	0.580
Satisfaction	CR	0.940	0.935
	AVE	0.612	0.591
n		769	769
Path Relationships			
Comprehensibility → Satisfaction		0.032	0.061*
Involvement → Satisfaction		0.166***	0.130**
Time Pressure → Satisfaction		-0.042	-0.112**
Trust → Satisfaction		0.448***	0.358***
Use → Satisfaction		0.216***	0.278***
R-squared		0.513	0.497

Notes. CR = composite reliability; AVE = average variance extracted; *Significance at 0.10; **Significance at 0.05; ***Significance at 0.001
Source. Authors' own calculations

Table A5. Structural path estimates for the binary label

Variables	Endogenous Construct	Path Coefficient	Standard Deviation	t-value	p-value
Comprehensibility	satisfaction	0.027	0.031	0.873	0.383
Involvement	satisfaction	0.166	0.038	4.421	0.000
Time pressure	satisfaction	-0.037	0.037	1.001	0.317
Trust	satisfaction	0.439	0.048	9.206	0.000
Use	satisfaction	0.227	0.051	4.486	0.000

Source. Authors' own calculations

Table A6. Structural path estimates for the multi-level label

Variables	Endogenous Construct	Path Coefficient	Standard Deviation	t-value	p-value
Comprehensibility	satisfaction	0.063	0.031	2.057	0.040
Involvement	satisfaction	0.127	0.044	2.908	0.004
Time pressure	satisfaction	-0.112	0.035	3.215	0.001
Trust	satisfaction	0.349	0.044	7.997	0.000
Use	satisfaction	0.291	0.049	5.919	0.000

Source. Authors' own calculations

