Business Strategy and the Environment
Bus. Strat. Env. **21**, 60–70 (2012)
Published online 10 July 2011 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/bse.722

Dynamic Adjustment of Eco-labeling Schemes and Consumer Choice – the Revision of the EU Energy Label as a Missed Opportunity?

Stefanie Lena Heinzle and Rolf Wüstenhagen*
Institute for Economy and the Environment, University of St Gallen, Switzerland

ABSTRACT
Products with a superior environmental performance, such as a high level of energy efficiency, are typically subject to information asymmetries. Therefore these product attributes are often undervalued in purchase decisions. Signaling, e.g. energy labeling, can overcome these asymmetries, with positive implications for effective consumer decisions, competitive advantage for suppliers of energy-efficient goods, and for societal goals such as mitigating climate change. However, there is a scarcity of research investigating how energy labels actually influence consumer choice. The recent revision of the European Union energy label provided a unique opportunity to investigate the effectiveness of energy labeling in a quasi field-experimental setting. We show that the proposed extension of the seven-point A–G rating scale by adding new classes A+, A++, etc. will result in a lower perceived importance of energy efficiency in consumer decision-making. Based on a stated preference survey investigating 2244 choices by German consumers, we conclude that the revision actually undermines the label's ability to overcome information asymmetries, hence potentially contributing to market failure. Copyright © 2011 John Wiley & Sons, Ltd and ERP Environment.

Received 2 March 2011; revised 14 April 2011; accepted 21 April 2011
Keywords: energy labeling; environmental policy; sustainable consumption; choice-based conjoint experiment; eco-innovation

Introduction

MARKETS FOR ENVIRONMENTALLY FRIENDLY PRODUCTS ARE CHARACTERIZED BY INFORMATION ASYMMETRIES between suppliers and consumers. For example, it is often difficult to identify the level of environmental performance of a product before purchasing it. In his seminal article on 'the market for lemons', Akerlof (1970) showed how the presence of information asymmetries can lead to market failure and adverse selection, and discussed signaling and screening as ways to overcome those challenges. One method of signaling that has received increasing attention from academics, policy makers and industry professionals is environmental or eco-labeling (Thogersen, 2000; De Boer, 2003; Pedersen and Neergaard, 2006; Rubik *et al.*, 2007). By providing information on the environmental performance of products, eco-labels can guide consumers towards a more environmentally friendly purchasing behavior (Grankvist and Biel, 2007).

*Correspondence to: Rolf Wüstenhagen, Good Energies Chair for Management of Renewable Energies, Institute for the Economy and the Environment, University of St Gallen, Tigerbergstrasse 2, 9000 St Gallen, Switzerland. E-mail: rolf.wuestenhagen@unisg.ch

Energy efficiency labels, i.e. third-party certification of products based on their energy consumption, are important examples of such eco-labels which play a significant role in reducing the energy consumption of appliances worldwide. Energy labels may be categorized as endorsement and comparison labels (Wiel and McMahon, 2005). Endorsement labels (e.g. the Energy Star issued by the US Environmental Protection Agency) follow a best-in-class approach in that they are only applied to the most energy-efficient products in a given category. Comparative labels (e.g. Australia's star rating scheme or the European Union energy label) rate the energy efficiency of a product in relation to an absolute scale (Harrington and Damnics, 2004). Those labels have become more widely applied in the field, and have induced technological improvements in the market for energy-efficient goods. However, there is a risk that they become a victim of their own success, especially in cases where there is no dynamic adjustment of labeling criteria. One such case is the European Union energy label: this rates the energy efficiency of products on a seven-point A to G scale, has been applied across a wide range of appliances since 1992 (European Commission, 1992, 2003) and has recently been adopted in a number of countries around the world including China, Brazil, Iran, Egypt and others. While the original idea was to only have the best products marked with an A rating, this highest energy efficiency class has become a de facto standard in many product categories, to an extent where up to 90% of products such as refrigerators, dishwashers and washing machines on the European market are now A-labeled (European Commission, 2010).

Two solutions were suggested to overcome this problem. One option, favored by consumer and environmental organizations (ANEC/BEUC, 2008), was to maintain the existing seven-point scale ranging from A to G, but tighten the criteria on a regular basis, so that every year only the most efficient products would be A-labeled. A product which would be placed at the top of the scale in one year would then be reclassified into a lower efficiency class in another year. This option would require the inclusion of a date on the label indicating how long the label would be valid (BUND/DUH, 2009). The other option, backed by industry associations (CECED, 2007), was to extend the scale by means of introducing new categories 'beyond A'. The energy efficiency class of one particular appliance would remain unchanged over time so that no updated sticker needed to be attached on the appliances in the store (ECEEE, 2009).

After months of negotiations and discussions of different proposals, in autumn 2009, members of the European Parliament and representatives from the European Commission and the EU Swedish Presidency finally reached the agreement to continue using letters A to G for classifications, but to allow introducing up to three additional classes on top of class A (A+, A++ and A+++ for the most efficient class) (Figure 1). However, the new proposal also limits the total number of energy classes to seven, i.e. if the highest energy class 'A+++' is introduced, the least energy efficiency class shown on the label will be a D (ECEEE, 2009).

The objective of the research presented in this paper is to assess if the proposed revision of the energy label by extending its scale is more effective in guiding consumer decisions towards energy-efficient goods than the proposal by consumer organizations to maintain the existing categories. To address this question, we investigated 2244 experimental choices made by 187 German consumers. We derive important implications for policy makers, managers and further research.

Figure 1. Illustration of energy efficiency classes of two label options ('A–G scale' versus 'A-plus scale')

DOI: 10.1002/bse

The Influence of Energy Labels on Consumer Decision Processes

Household appliances are a major source of energy demand and represent one of the fastest growing energy loads (OECD/IEA, 2003). The long-term increase in household electricity use is often associated with a growing ownership of electrical appliances (Bertoldi and Atanasiu, 2009). By switching to the best technologies on the market, a huge potential saving in electricity consumption could be achieved (IEA, 2007). However, although energy-efficient products are typically characterized by a lower life-cycle cost than conventional products (Känzig and Wüstenhagen 2010), energy-efficient investments that appear to be cost-effective on an estimated life-cycle cost basis are often not made (Ruderman *et al.* 1987), a phenomenon known in the literature as the 'energy efficiency gap'. One reason for this energy efficiency gap is that consumers are often not aware of the fact that the appliance they are about to purchase is also an energy service having running costs such as costs for electricity etc. (Wilkenfeld *et al.*, 1998).

Energy efficiency labels are particularly well suited to narrowing the energy efficiency gap and mitigating potential inefficiencies resulting from imperfect information distribution about energy use. Energy labels can be used to provide information to consumers and to enable them to compare the energy efficiency of appliances on an equitable basis by acting as an indicator showing how energy efficient the product is (Mahlia *et al.*, 2002). Furthermore, such labels help manufacturers to gain a competitive advantage by producing environmentally friendly products (Thogersen, 2000) and therefore act as an especially effective and essential element in any government's portfolio of energy-efficient policies (Wiel and McMahon, 2005).

The effectiveness of energy labels can only be evaluated based on an understanding of consumer behavior. However, empirical data regarding the impact of energy labels on consumers' responses are unusually limited and there is a huge lack of studies on how consumers recognize, perceive, understand and consider the information on the labels in their purchasing decisions. In one of the few studies available, Sammer and Wüstenhagen (2006) examined the impact of the EU energy label on consumer choice among different washing machines with different degrees of energy efficiency and found that Swiss consumers are willing to pay a premium of 347 Swiss Francs for a washing machine with an energy efficiency level of A in comparison to a washing machine with a B rating. Furthermore, Shen and Saijo (2009) conducted a choice experiment in China to examine whether the Chinese energy efficiency label influences consumers' choices for air conditioners and refrigerators. The respondents' awareness of the label was rather high and it was found that the energy efficiency classes significantly influenced the choice of air conditioners and refrigerators. Ward *et al.* (2011) conducted a choice experiment to gather data on consumers' willingness to pay (WTP) for an Energy Star label on a refrigerator in the United States. The survey suggests that consumers' WTP amounts to an extra US$249.82 to US$349.30 for a refrigerator that has been awarded the Energy Star label. Finally, Saidur *et al.* (2005) investigated the best format for an energy guide label for household refrigerator–freezers in Malaysia. Different labeling concepts of other countries were tested, and the authors concluded that the star labeling worked best with the majority of respondents.

However, no study to date has investigated the impact of a modification of a labeling scheme by adding new classes on top of the current highest class. The recent revision of the European Union energy label provided a unique opportunity to investigate this effect. We aim to provide a more nuanced understanding of how the introduction of new categories influences consumer behavior, and what conclusions can be derived for the design of effective labeling schemes.

As the new energy label with the new format and the additional classes A+, A++ and A+++ has not been fully introduced for televisions yet, no market data are currently available about revealed preferences. An important requirement for using a revealed preference approach is that sufficiently long data series are available. Thus, for our study it was not possible to observe people's actual purchasing decisions. Accordingly, an appropriate methodological approach was necessary to measure stated preferences. In contrast to the revealed preferences approach, which observes actual choices made by decision-makers in real market circumstances, stated preferences are derived from preferred choices made under different hypothetical scenarios in experimental markets (Danielis and Rotaris, 1999).

Methodological Considerations

Particularly in the area of individual decision-making behavior in marketing, but also in the field of environmental economics, the stated preference approach using choice experiments is widely applied (Train, 2003, Hensher

DOI: 10.1002/bse

et al., 2005). This study also makes use of a stated preference choice experiment, a choice-based conjoint (CBC) experiment, in which consumers are given a hypothetical setting where they are asked to choose their preferred alternative described by a set of criteria from numerous presented sets. Thus, by forcing consumers to decide which characteristics are most important, they have to make tradeoffs between different levels of product attributes. By conducting a choice experiment, it is possible to measure preferences in simulated quasi-realistic decision/purchasing situations since the decision-making criteria are not presented separately, but simultaneously (Orme, 2006; Lilien *et al.*, 2007). Briefly described, a CBC experiment considers a quasi-realistic buying situation, where consumers choose between one or more products from a restricted product set. By choosing the most beneficial product from this restricted set, the preferences of the respondents can be directly derived (McFadden, 1974).

Whereas fridges and freezers, washing machines and dishwashers have been labeled for more than a decade, televisions have not been part of the European Union labeling scheme up to now. Within the last couple of years, the television market has experienced an ongoing trend towards increasingly larger screen sizes, which has resulted in very high power consumption during viewing times (GfK, 2008). Televisions can therefore be classified as high-energy consuming appliances and consequently there is a large energy reduction potential in adding this category to the European energy labeling scheme; this is what makes television sets an interesting product category for this research.

We used a between-subjects design where two different independent experimental groups were surveyed. Respondents were split in two different samples (sample 1 'A–G closed' scale format; sample 2 'A-plus' scale format), which only differed with regard to the presentation format of the label. Technically, the set of attributes and levels for both subgroups was identical. Therefore, differences in the preference structure between the two subgroups could be traced back to the different versions of the label. We did not make explicit assumptions about the underlying distribution of products on the market across the seven energy efficiency classes in both subsamples, which implicitly suggests a similar distribution of products from A to G in sample 1 and from A+++ to D in sample 2. Given the skewed distribution in today's market, where products have become crowded at the top end of the scale, our findings from sample 1 can consequently only be generalized if the current scale is subject to a dynamic adjustment of the criteria, so that a reasonably even distribution will be restored. In line with methodological state-of-the-art in conjoint analysis, we used an experimental design with the same number of levels for each attribute, namely four levels. We chose to include the four highest classes for each label version (A, B, C and D for sample 1 and A+++, A++, A+ and A for sample 2).

The first stage in the design of the study involved the identification of the most important product attributes and their levels for televisions. In order to select decision-relevant product categories we conducted expert interviews (e.g. with retailers) and reviewed marketing documents (e.g. catalogs, websites). The attributes and the attribute levels that were presented in the choice tasks are listed in Table 1; a typical choice task is displayed in Figure 2. The chosen brands and equipment versions represent a spectrum of the German market for televisions. The realistic price range we chose represents a continuum from low to high prices of comparable TV sets usually available in Germany. For the attribute levels of the energy label, we chose to include the four highest classes for both versions of the label as described above. We decided not to include the attributes size and technology (e.g. plasma, LCD) in order to guarantee the independence of the attributes from each other and to avoid unrealistic bundles of attribute levels due to random combination.

We used Sawtooth software, the standard application for conjoint analysis in marketing research, to develop a computer generated, choice-based, conjoint design. The choice tasks were randomly calculated. We applied a full profile method, i.e. all attributes were presented for each set of alternatives. The randomized design accounted for the design principles of minimal overlap, level balance and orthogonality (Huber and Zwerina, 1996). All respondents received a series of 12 choice tasks involving comparisons of different televisions with varying levels of attributes. Each choice task presented four different television alternatives where respondents had to choose their preferred alternative.

Sample Characteristics

This study is based on 2244 choice observations in Germany, based on 12 choices each of 187 consumers. Sample 1 (label version 'A–G scale') includes 1080 choice tasks and sample 2 (label version 'A-plus' scale) is based on data for 1164 choice tasks. These respondents were recruited by a professional marketing research company (GfK), who

DOI: 10.1002/bse

Attributes	Attribute levels	
	Sample 1 (‘A–G’ scale)	Sample 2 (‘A-plus’ scale)
Brand	Samsung	Samsung
	Sony	Sony
	Philips	Philips
	TCM of Tchibo	TCM of Tchibo
Equipment version[a]	Simple*	Simple*
	Medium**	Medium**
	High-tech***	High-tech***
Energy label	A	A+++
	B	A++
	C	A+
	D	A
Purchase price	€499	€499
	€649	€649
	€799	€799
	€949	€949

Table 1. Attributes and attribute levels in the choice tasks
[a]Equipment version:
*Simple: HD ready, 1× HDMI, response time 8, contrast ratio 5000:1.
**Medium: HD ready, 2× HDMI, USB, response time 6, contrast ratio 10,000:1.
***High-tech: Full HD, 4× HDMI, PC connection, USB, response time 4, contrast ratio 50,000:1.

conducted computer-assisted personal interviews (CAPI) in 2009. The target population of the study consisted of the general German population. The sample was drawn by quota sampling, taking into account the distribution of the target population by state (German Bundesland), city size, household size, and sex. Setting quotas using these indicators is a standard procedure for drawing representative samples in professional market research. As evidenced by chi-square tests, no significant differences were found between the two samples with regard to city size ($\chi^2 = 0.268$, 3 d.f., $P = 0.966$), household size ($\chi^2 = 0.423$, d.f. 2, $P = 0.809$), sex ($\chi^2 = 0.124$, d.f. 1, $P = 0.725$) and region within Germany[1] ($\chi^2 = 2.466$, d.f. 3, $P = 0.482$). We then compared the characteristics of both samples with national data from the German Federal Statistics Office (2009). Both samples were compared with the German population using a chi-square test and no statistically significant difference was found. Sample 1 was not significantly different from the German population with regard to city size ($\chi^2 = 3.820$, d.f. 3, $P = 0.282$), household size ($\chi^2 = 4.324$, d.f. 2, $P = 0.115$), sex ($\chi^2 = 0.036$, d.f. 1, $P = 0.849$) and region within Germany ($\chi^2 = 1.415$, d.f. 3, $P = 0.702$). The same was true for Sample 2 which was not significantly different from the German population with regard to city size ($\chi^2 = 5.554$, d.f. 3, $P = 0.135$), household size ($\chi^2 = 4.515$, d.f. 2, $P = 0.105$), sex ($\chi^2 = 0.097$, d.f. 2, $P = 0.756$) and region within Germany ($\chi^2 = 3.885$, d. f. 3, $P = 0.274$). Table 2 shows in detail how the two subsamples compare with the overall population.

Results

By using hierarchical Bayesian (HB) estimation, utilities at the individual level can be estimated (Rossi and Allenby, 2003). HB analysis is regarded as being a state-of-the-art method for estimating utilities from CBC studies. Compared with traditional aggregate models (e.g. multinomial logit analysis) the HB approach

[1]In order to calculate a chi-square statistic, the cells of a chi-square contingency table must contain more than five cases. As several German states did not contain more than five cases, we had to form four regional groups: Group 1: northern region (Bremen, Hamburg, Lower Saxony, Schleswig-Holstein); Group 2: eastern region (Mecklenburg-Western Pomerania, Brandenburg, Berlin, Saxony, Saxony-Anhalt, Thuringia); Group 3: western region (North Rhine-Westphalia, Hesse, Rhineland-Palatinate, Saarland); Group 4: southern region (Bavaria, Baden-Württemberg).

DOI: 10.1002/bse

The European Union is planning to introduce a new label for televisions, which will look like the following:

The colour "green" indicates low energy consumption, the colour "red" indicates very high consumption of energy. If these were your only options, which one would you choose? Choose by clicking one of the buttons below:

Brand	Philips	Samsung	Sony	TCM of Tchibo
Equipment version	High-Tech***	Medium **	Medium **	Simple*
Energy efficiency class	A	B	C	D
Price	949€	799€	649€	499€
	○	○	○	○

Equipment version:
* Simple: HD-Ready, 1x HDMI, response time 8, contrast ratio 5000:1
** Medium: HD-Ready, 2x HDMI, USB, response time 6, contrast ratio 10000:1
*** High-tech: Full-HD, 4x HDMI, PC connection, USB, response time 4, contrast ratio 50000:1

The European Union is planning to introduce a new label for televisions, which will look like the following:

The colour "green" indicates low energy consumption, the colour "red" indicates very high consumption of energy. If these were your only options, which one would you choose? Choose by clicking one of the buttons below:

Brand	Philips	Samsung	Sony	TCM of Tchibo
Equipment version	High-Tech***	Medium **	Medium **	Simple*
Energy efficiency class	A+++	A++	A+	A
Price	949€	799€	649€	499€
	○	○	○	○

Equipment version:
* Simple: HD-Ready, 1x HDMI, response time 8, contrast ratio 5000:1
** Medium: HD-Ready, 2x HDMI, USB, response time 6, contrast ratio 10000:1
*** High-tech: Full-HD, 4x HDMI, PC connection, USB, response time 4, contrast ratio 50000:1

Figure 2. Sample choice task for two samples

significantly improves the analysis of preferences. While earlier methods combined data for all individuals and were criticized for obscuring important aspects of the data, with a Bayesian framework it is possible to analyze choice data at the individual level (see Rossi and Allenby, 2003, and Huber and Train, 2001, for more detailed discussion of hierarchical modeling). Table 2 presents the average utilities of each attribute level of the HB model for televisions where the raw part-worth utilities were rescaled by a method called zero-centered Diffs. The Diffs method rescales utilities so that the total sum of the utility differences between the worst and best levels of each attribute across attributes is equal to the number of attributes times 100 (Sawtooth Software, 1999). The results of the hierarchical Bayes model are presented in Table 3.

With regard to data quality, the average root likelihood (RLH) can be used as a measure of fit to assess convergence of HB estimates. RLH is the geometric mean of the predicted probabilities (Sawtooth Software, 2009). In this study, as each choice task presented four alternatives, it would be predicted that each alternative would be chosen with a probability of 25% (corresponding RLH of 0.25). RLH was 0.648 for the model of sample 1 ('A–G' scale) and 0.675 for the model of sample 2 ('A-plus' scale format). The relatively large values indicate good fit of the two overall models. The actual values of 0.648 for sample 1 and 0.675 for sample 2 indicate that these iterations were about 2.6 or 2.7 better than the chance level.

DOI: 10.1002/bse

Characteristics	'A–G closed' scale format		'A-plus' scale format		German population[a]
	n	%	*n*	%	
	90	100%	97	100%	100%
Region					
Northern region (Bremen, Hamburg, Lower Saxony, Schleswig-Holstein)	18	20.0%	13	13.4%	16.1%
Eastern region (Mecklenburg-Western Pomerania, Brandenburg, Berlin, Saxony, Saxony-Anhalt, Thuringia)	19	21.1%	27	27.8%	20.1%
Western region (North Rhine-Westphalia, Hesse, Rhineland-Palatinate, Saarland)	31	34.4%	30	30.9%	35.4%
Southern region (Bavaria, Baden-Württemberg)	22	24.5%	27	27.8%	28.4%
City size					
$n = 1–19{,}999$	39	43.3%	32	42.3%	41.7%
$n = 20{,}000–99{,}999$	26	28.9%	25	26.8%	27.4%
$n = 100{,}000{,}499{,}999$	17	18.9%	21	21.6%	15.0%
$n > 500{,}000$	8	8.9%	9	9.3%	15.9%
Household size (persons)					
1	27	30.0%	31	28.9%	39.4%
2	39	43.3%	34	40.2%	34.0%
3 or more	24	26.7%	22	30.9%	26.6%
Sex					
Female	45	50.0%	38	47.4%	51.0%
Male	45	50.0%	49	52.6%	49.0%

Table 2. Description of sample characteristics
[a]German Federal Statistics Office (2009).

Importances

In a following step, conjoint importances were computed. Importances describe how much influence each attribute has on the purchase decision. The importance of attributes in influencing the purchasing decision can be measured by comparing the difference between the highest and lowest part-worth utility of its levels. Conjoint importances are displayed in Table 4.

In both samples, the most important product attribute of a television was the purchase price, followed by the energy label, the equipment version and the brand. However, there were differences in conjoint importances of the attribute energy label between sample 1, with 33.6%, and sample 2, with 23.0%. This analysis shows that an energy label with an 'A–G scale' has over 10 percentage points more influence than an energy label with an 'A plus' scale. Introducing the new label with its additional categories (A+, A++, A+++) weakened the efficacy of the label, resulting in lower consumer awareness of energy efficiency as an important attribute. The statistical differences between groups for the importance level of the attribute energy label were compared using a non-parametric test, that is, the Mann–Whitney U-test ($P > 0.05$, two-sided). Significant differences were found between sample 1 and sample 2 ($P < 0.001$). The change in the energy label's importance due to its new design led to an increase in importance of one or more other attributes, *ceteris paribus*. Thus, consumers switched away from energy-efficient products and focused on other product characteristics during their purchase decision.

Simulation of Market Response

A market simulator can be used to convert individual part-worths from HB estimation into simulated market choices and to compute shares of preferences for competing product alternatives. Market simulation models are used to analyze consumer choices for a defined set of products and their specific product features. Share of preference can be defined as the percentage of respondents who would prefer one of the specified products. For our

Attribute level	Sample 1 ('A–G closed' scale format) $n=90$ Average utilities	(SD)	Sample 2 ('A-plus' scale format) $n=97$ Average utilities	(SD)
Brand				
Samsung	6.01	(24.36)	1.00	(14.56)
Sony	3.42	(15.30)	7.82	(17.34)
Philips	8.27	(21.47)	8.91	(17.67)
TCM of Tchibo	-17.70	(29.62)	-17.73	(22.70)
Equipment version[a]				
Simple	-28.46	(36.39)	-47.14	(32.25)
Medium	-1.62	(18.39)	3.24	(13.10)
High-tech	30.08	(36.74)	43.89	(33.89)
Energy label				
A/A+++	61.82	(48.01)	39.51	(30.67)
B/A++	23.49	(24.27)	19.05	(13.90)
C/A+	-21.65	(27.53)	-9.82	(13.99)
D/A	-63.66	(34.89)	-48.74	(29.19)
Purchase price				
€499	61.18	(54.72)	83.55	(51.77)
€649	25.04	(26.75)	23.88	(20.60)
€799	-25.30	(23.68)	-25.88	(32.24)
€949	-60.92	(41.63)	-81.55	(26.94)

Table 3. Results of the hierarchical Bayes model for televisions
[a]See note to Table 1 for explanation of Equipment version.

Attributes	Sample 1 ('A–G' scale format)	Sample 2 ('A-plus' scale format)
Brand	13.4%	10.9%
Equipment version	18.6%	23.6%
Energy label	33.6%	23.0%
Purchase price	34.5%	42.6%

Table 4. Relative attribute importances derived from hierarchical Bayes estimation of utilities

analysis, we applied a randomized first choice simulation method to estimate share of preference. A 'maximum utility rule' is assumed, which predicts that respondents would choose the option with the highest composite utility. Randomized first choice simulations then estimate the choices of each participant, adding random error to the utility values at each of 100,000 iterations and averaging those predictions across iterations and respondents (see Huber *et al.*, 1999, and Orme, 2006, for more detailed discussions of the computation of randomized first choice simulations).

In the following scenario, a realistic market situation was demonstrated by calculating the share of preference of four hypothetical products. Reflecting the real market situation, the price of the appliance varied according to the energy efficiency class (i.e. the most expensive television came with the highest energy efficiency class, whereas the cheapest television was labelled with the lowest energy efficiency class). The attributes brand and equipment were set at a constant level to allow testing of the isolated effect of the combination of energy efficiency class and price.

The results in Table 5 show that respondents of sample 1 ('closed A–G' scale format) were about 4.5 times more likely to choose the television with the highest energy efficiency class in combination with the highest price than respondents from sample 2 ('A-plus' scale format) (33.7% vs. 7.5%). Respondents of sample 1 were about 1.5 times less likely to choose the television with the lowest energy efficiency class in combination with the lowest price than respondents from sample 2 (30.8% vs. 45.7%). Thus, in sample 1, the preference share for the television with the highest energy efficiency class in combination with the highest price was about 2.9% higher than the preference

DOI: 10.1002/bse

Attributes	Highest energy efficiency class and highest price		Second highest energy efficiency class and second highest price		Second lowest energy efficiency class and second lowest price		Lowest energy efficiency class and lowest price	
Sample	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
Brand	Samsung		Samsung		Samsung		Samsung	
Equipment version[a]	Medium		Medium		Medium		Medium	
Energy label	A	A+++	B	A++	C	A+	D	A
Price	€949		€799		€649		€499	
SoP (%)	33.7%	7.5%	19.1%	21.4%	16.5%	25.4%	30.8%	45.7%
Standard error	4.0	2.1	2.6	3.1	2.8	3.2	4.1	4.3

Table 5. Share of preference (SoP) of four hypothetical products
[a]See note to Table 1 for explanation of 'Medium' equipment version.

share for the television with the lowest energy efficiency class in combination with the lowest price. In contrast, the preference share in sample 2 for the high-efficiency, but expensive television was 38.2% lower than the preference share for the low-efficiency, but cheap, television set.

The statistical differences between the two samples were compared using a Mann–Whitney U-test ($P > 0.05$, two-sided). Significant differences were found between samples 1 and 2 ($P < 0.001$). We can therefore conclude that an increase from a D to an A labelled television produces enough utility for respondents in sample 1 so that the shares of preference are more than equalized although the price goes up. In other words, respondents of sample 1 are willing to put up with a high price if the energy efficiency class is high. Our analysis therefore proves that respondents of sample 1 have a higher willingness to pay for energy-efficient appliances than respondents of sample 2.

Conclusions

This study showed how the effectiveness of a well-established energy labeling scheme can actually be diminished by the introduction of new rating categories.

Our research has important implications for policy makers. The fact that the effectiveness of the European energy label decreases with the introduction of new categories beyond A illustrates that labels and brands, which intend to reduce complexity for consumers, operate under narrow constraints. Labels can reduce uncertainty and overcome information asymmetry, but in order to do so, they need to present consumers with a meaningful reduction of complexity. Going from a closed scheme to an extended scheme adding new categories, the effectiveness of the well-established label is reduced. The results clearly show that introducing the new label with its additional categories weakens the effect of the label, resulting in lower consumer awareness about energy efficiency as an important attribute. Policy makers can conclude from our study that responding to industry requests for 'more flexibility' can result in more complexity for consumers and actually countervail their efforts to increase consumer awareness about the real energy use of appliances.

Given that the new labeling scheme was the result of a political compromise and was strongly backed by industry associations, a question arises about the effectiveness of participatory decision-making in environmental policy, and especially the role of firms' and industry associations' non-market strategies. While it remains an interesting area for further research to explore why the European industry associations actually backed the 'A plus' scale, our findings tend to suggest that their stance in the political negotiations may actually not have been in the best interest of those manufacturers who showed technological leadership and could have maintained a competitive advantage from the current clear labeling scheme, combined with dynamic adjustment of the criteria for reaching the A rating. Our analysis shows that the impact of an 'A–G scale' on consumers' decisions is much stronger and

DOI: 10.1002/bse

therefore consumers are more willing to pay a higher premium for the highest classes of the 'A–G scale' than for the classes of the 'A-plus' scale. This strong WTP for a labeled product should have been encouraging for manufacturers to support the maintenance of the well-known A–G scheme in order to differentiate themselves based on energy-efficient products. By reaping the benefit of this higher latent WTP, manufacturers who show technological leadership might get a higher return on their investment in R&D with the 'A–G scale' scheme.

The implications of our research may not be limited to the context we investigated, the effectiveness of energy labeling, but apply to a broader field of applications. Two areas that seem to have particular similarities are the influence of rating scales commonly used in product reviews (e.g. on travel platforms, movie or book ratings, etc.) on consumer behavior and the influence of credit ratings on investor behavior. Given the large influence that rating scales have on consumers' purchase decisions in different choice contexts, it is critical to understand how an alteration of the scale levels would influence the perception of such rating scales. As for credit ratings, the recent financial crisis has highlighted the critical role of rating agencies in guiding investor decision-making. Quite similar to the extension of the European energy labeling scheme that we investigated in this paper, such ratings have become ever more fine-grained over time, and it seems worth studying whether investor perceptions of, for example, the difference between an AA– and a BBB+ rating is congruent with the underlying risk differences that rating agencies intend to signal. Again, it may be fruitful to explore potential asymmetries in investor reactions to such ratings. While it goes beyond the scope of our paper to speculate about the findings of such further research, it would certainly be interesting to see whether it could derive similarly clear conclusions for the future design of effective credit ratings.

In conclusion, findings from this research provide a rich source of information to guide future research that focuses on the influence of eco-labeling on consumer decision-making. With regard to further research streams, applying a choice-based conjoint analysis to analyse the effect of the introduction of additional energy classes on other product categories (e.g. washing machines, refrigerators, etc.) would provide interesting research opportunities. Also, as we conducted our research only in Germany, comparing the influence of labeling schemes across different countries would be fruitful.

Acknowledgements

This paper is based on research funded by the German Ministry of Education and Research, within the program 'Socio-Ecological Research (SÖF)', project 'Social, environmental and economic dimensions of sustainable energy consumption in residential buildings (seco@home)', contract no. 01UV0710, coordinated by Dr Klaus Rennings at the Center for Economic Research (ZEW), Mannheim. The authors wish to acknowledge valuable support and feedback from the seco@home project team and advisory board. We also thank the editor, two anonymous reviewers, Dr. Martin Meissner, Nina Hampl and Dr. Moritz Loock for their comments. All remaining errors are the sole responsibility of the authors.

References

Akerlof GA. 1970. The market for 'Lemons': quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **84**(3): 488–500.

ANEC/BEUC. 2008. *Consumers strongly in favour of keeping the A-G Energy Label*. http://www.anec.eu/attachments/ANEC-PR-2008-PRL-009.pdf [28 February 2011].

Bertoldi P, Atanasiu B. 2009. *Electricity consumption and efficiency trends in European Union - Status report 2009*. JRC Scientific and Technical Reports, Ispra, Italy.

BUND/DUH. 2009. *Stellungnahme zur Neugestaltung der EU-Energieverbrauchskennzeichnung*. http://www.bund.net/fileadmin/bundnet/pdfs/klima_und_energie/20091111_energie_label_stellungnahme.pdf [28 February 2011].

CECED. 2007. *"Beyond A" - a dynamic new labelling scheme is necessary for promoting continuous improvement of the energy performance of home appliances*. European Committee of Domestic Equipment Manufacturers, Brussels.

Danielis R, Rotaris L. 1999. Analysing freight transport demand using stated preference data: a survey. *Transporti Europei* **13**: 30–38.

De Boer J. 2003. Sustainability labeling schemes: the logic of their claims and their function for stakeholders. *Business Strategy and the Environment* **12**(4), 254–264.

ECEEE. 2009. *The Energy Labelling Directive*. http://www.eceee.org/Energy_Labeling [28 February 2011].

DOI: 10.1002/bse

European Commission. 1992. Council directive 92/75/EEC of 22 September 1992 on the indication by labelling and standard product information of the consumption of energy and other resources by household appliances. European Commission, Brussels.
European Commission. 2003. Commission Directive 2003/66/EC of 3 July 2003 amending Directive 94/2/EC implementing Council Directive 92/75/EEC with regard to energy labelling of household electric refrigerators, freezers and their combinations. European Commission, Brussels.
European Commission. 2010. Questions & Answers: new energy labels for televisions, refrigerators, dishwashers and washing machines. MEMO/10/451. European Commission, Brussels.
German Federal Statistics Office. 2009. *Statistisches Jahrbuch. Kapitel Bevölkerung*. Statistisches Bundesamt (Federal Statistics Office), Wiesbaden.
GfK. 2008. *How to combine consumer electronics and environment?* http://www.gfk.com/group/press_information/press_releases/002598/index.en.html [28 February].
Grankvist G, Biel A. 2007. The impact of environmental information on professional purchasers' choice of products. *Business Strategy and the Environment* **16**(6): 421–429.
Harrington L, Damnics M. 2004. *Energy Labelling and Standards Programs throughout the World*. Edition 2.0. The National Appliance and Equipment Energy Efficiency Committee, Victoria.
Hensher DA, Rose JM, Greene H. 2005. *Applied Choice Analysis – A primer*. University Press, Cambridge.
Huber J, Orme BK, Miller R. 1999. Dealing with product similarity in conjoint simulations. Sawtooth Software Conference Proceedings.
Huber J, Train K. 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. *Marketing Letters* **12**: 259–269.
Huber J, Zwerina K. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing Research* **33**(3): 307–317.
IEA. 2007. *Energy efficiency policy recommendations to the G8 2007 summit*. http://www.iea.org/g8/docs/final_recommendations_heiligendamm.pdf [28 February].
Känzig J, Wüstenhagen R. 2010. The effect of life cycle cost information on consumer investment decisions regarding eco-innovation. *Journal of Industrial Ecology* **14**(1): 121–136.
Lilien GL, Rangaswamy A, De Bruyn A. 2007. *Principles of Marketing Engineering*. Bloomington, Trafford.
Mahlia TMI, Masjuki HH, Choudhury IA. 2002. Potential electricity savings by implementing energy labels for room air conditioner in Malaysia. *Energy Conversion and Management* **43**(16): 2225–2233.
McFadden D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, Zarembka P (ed.). Academic Press, New York, 105–142.
OECD/IEA. 2003. Cool appliances – policy strategies for energy efficient homes. Energy efficient policy profiles, OECD/IEA, Paris.
Orme BK. 2006. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Research Publishers, Madison.
Pedersen ER, Neergaard P. 2006. Caveat emptor – let the buyer beware! Environmental labeling and the limitations of 'green' consumerism. *Business Strategy and the Environment* **15**(1): 15–29.
Rossi PE, Allenby GM. 2003. Bayesian statistics and marketing. *Marketing Science* **22**: 304–328.
Rubik F, Frankl P, Pietroni L, Scheer D. 2007. Eco-labelling and consumers: towards a re-focus and integrated approaches. *International Journal of Innovation and Sustainable Development* **2**(2): 175–191.
Ruderman H, Levine M, McMahon J. 1987. Energy-efficiency choice in the purchase of residential appliances. In *Energy Efficiency: Perspectives on Individual Behavior*, Kempton W, Neiman M (eds.). American Council for an Energy-Efficient Economy, Washington, 41–50.
Saidur R, Masjuki HH, Mahlia TMI. 2005. Labeling design effort for household refrigerator-freezers in Malaysia. *Energy Policy* **33**(5): 611–618.
Sammer K, Wüstenhagen R. 2006. The influence of eco-labelling on consumer behaviour – Results of a discrete choice analysis for washing machines. *Business Strategy and the Environment* **15**(3): 185–199.
Shen J, Saijo T. 2009. Does an energy efficiency label alter consumers' purchasing decisions? A latent class approach on a stated choice experiment in Shanghai. *Journal of Environmental Management* **90**(11), 3561–3573.
Sawtooth Software. 1999. Scaling conjoint part worths: Points vs. zero-centered diffs. http://www.sawtoothsoftware.com/education/ss/ss10.shtml [28 February].
Sawtooth Software. 2009. The CBC/HB system for hierarchical Bayes estimation. Version 5.0 technical paper. Technical Paper Series.
Thogersen J. 2000. Psychological determinants of paying attention to eco-labels in purchase decisions: model development and multinational validation. *Journal of Consumer Policy* **23**(3): 285–313.
Train K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.
Ward DO, Clark CD, Jensen KL, Yen ST, Russell CS. 2011. Factors influencing willingness to pay for the Energy Star label. *Energy Policy* **39**(3): 1450–1458.
Wiel S, McMahon JE. 2005. Energy-efficiency labels and standards: a guidebook for appliances, equipment and lighting. 2nd edition. Collaborative Labeling and Appliance Standards Program (CLASP), Washington D.C.
Wilkenfeld G, Waide P, Du Pont P. 1998. International update on the status of energy labelling and MEPs: Part 1 – Energy labelling of household appliances. Draft report prepared for National Appliance and Equipment Energy Efficiency Committee.

DOI: 10.1002/bse