Investigating the effect of introducing a tier to a multi-tiered ecolabel on consumer willingness to pay for sustainable products.

Master's Thesis by Frederic Pieter Caspar Klapwijk (2651102)

MSc Spatial, Transportation and Environmental Economics (STREEM)

Supervisor: dr. ir. Erik Ansink

Abstract

The effect of multi-tiered traffic light ecolabels is not yet completely understood. To gain insight in this type of ecolabels this study tests the hypothesis that increasing the number of tiers of a multi-tiered traffic light ecolabel will have a positive effect on willingness to pay (WTP) for the highest tiered label. To better understand the mechanisms of traffic light ecolabels, a second hypothesis, that price is a proxy for environmental quality during information scarcity, was tested. To test these hypotheses, a discrete choice experiment was conducted with two treatments, using a hypothetical ecolabel with two- and three tiers. Mixed logit was used to analyze the experiment data and estimate willingness to pay. Though willingness to pay for the highest tiered label was found to be significant and positive, no evidence was found that willingness to pay differed between treatments. Some evidence was found that price can be a proxy for environmental quality during information scarcity.

Contents

1. Introduction[1]

One of the tools that can be used to combat climate change is nudging consumer behavior, using information to turn at least some consumers away from the products with the highest environmental impact. Because of this kind of information people are starting to change their behavior, such as taking the train for medium distances, instead of flying. Information provisioning is complementary to numerous other tools in a policymakers toolbelt, such as emission taxing or investment in renewable energy sources, which are not topic of this study.

One of the easiest ways to convey information about the environmental impact of a product is by putting environmental labels on product packaging. There is currently no one label that conveys environmental impact for all consumer goods. There are over 450 ecolabels worldwide(***All ecolabels | Ecolabel Index***.), fragmenting the information. This fragmentation is not optimal, as effectiveness of ecolabels largely depends on familiarity with the label (Hainmueller et al., 2015; Sogn-Grundvåg et al., 2014; Vlaeminck et al., 2014). With so many different labels in circulation, it is becoming increasingly difficult for consumers to keep track of what each of them stands for, and what the requirements are for certification. It is no wonder that label familiarity is low, at least in Europe (Brécard, 2014). What makes it worse is that most of these ecolabels are binary labels. Binary labels usually have only one set of requirements to comply with to be certified. They simultaneously have the potential to contain a large amount of data, and too little. For the investigative consumer the label can convey a slew of information through the requirements, but that takes time and effort to research. Binary labels are effective ways to communicate certain product characteristics to consumers (Hainmueller et al., 2015; Sogn-

[1] Partially based on my research project (Frederic Pieter Caspar Klapwijk, 2020)

Grundvåg et al., 2014; Xu et al., 2012), but they do not provide an environmentally conscious consumer with a comparison between products with the same label.

Traffic light ecolabels are one solution for this problem, providing more information at a glance. Traffic light labels are a subset of multi-tiered labels, where the label color conveys the tier. With one look at the label, a consumer can see how the product they are looking at is rated and can easily compare it to other products with the same label. The EU energy usage label is a prime example, where with a color scheme, differently sized bars, and a letter, you get a lot of information instantaneously. This information can potentially influence purchasing decisions, and, if implemented correctly, nudge consumers towards buying more energy efficient products (Stadelmann & Schubert, 2018).

This study will investigate the effect of conveying more information about the environmental impact of a product, through the use of an ecolabel. My expectation is that increasing the number of tiers of a multi-tiered traffic light ecolabel will have a positive effect on willingness to pay (WTP) for the highest tiered label. To test this hypothesis, a discrete choice experiment was designed with two treatments, based on two different ecolabels: a two-tiered label, either red or green, and a three-tiered label with red, yellow, and green tiers. These labels are versions of the hypothetical enviro-score label, a color-coded multi-tier label with a summarized letter grade, summarizing all environmental effects of the product. Comparing the two treatments, if WTP for the highest tier in the three-tiered label treatment is significantly greater than in the two-tiered label treatment the hypothesis is confirmed. If not, then my hypothesis, that increasing the number of tiers of a multi-tiered traffic light ecolabel will have a positive effect on willingness to pay (WTP) for the highest tiered label, must be rejected.

In the two-tiered choice experiment the middle tier is replaced with one of the other tiers. For

this situation, a second hypothesis is posed that tests one possible mechanism for different outcomes under the two treatments, if any; when products appear to be homogeneous due to information scarcity, consumers use price as a proxy for environmental quality.

The following section will give an overview of previous research regarding the effects of traffic light label schemes. Section 3 describes the design of the choice experiment and data collection. Section 4 then introduces the methodology to analyze the data. Section 5 and 6 provide an overview of the results and discuss them, respectively. Finally, Section 7 concludes.

2. Literature review

Teisl et al., (2008) have created a multi-equation psycho-economic model of consumer reactions to ecolabelling, attempting to capture the effects that affect consumer decision making when purchasing new cars. They found that the lack of information regarding the emissions of vehicles in the marketplace makes consumers less likely to purchase a more environmentally friendly car. Over the past decade this information has certainly increased in various ways, but the effect of information should still be relevant.

One example of a traffic light label actually in use is the nutri-score, a French nutritional label in the process of being adopted in several European countries. It has been found to be an effective method to convey information to consumers during their purchasing decision (Chantal & Hercberg, 2017). In studying this nutri-score, Egnell et al., (2018) found that traffic light labels increases consumers' ability to identify healthier products. Labels with a summarized grade and minimal further information performed significantly better at this than labels which conveyed additional information, such as exact sugar content.

Similar results were found using color-coded labels in a supermarket setting. These labels were found to be effective to shift consumption from environmentally harmful food products towards environmentally friendly ones (Vlaeminck et al., 2014). The label used in this study did convey a lot of additional information, possibly making it less effective than a summarized label.

Thorndike et al., (2014) studied the same effect in a busy hospital cafeteria for a period of 24 months. They found that a three-tiered labeling scheme caused overall sales for both red and yellow products to drop significantly after introduction of the label by 20% and 4% respectively, coupled with an increase in sales of green products by 12%. With beverages this effect was different, the sales of red beverages dropped by 39%, whereas yellow and green increased by

16% and 10%, respectively. These results indicate that the effect of the traffic light is not homogeneous between different product groups. This could also mean that the cafeteria supply of healthy products, before introduction of the label, was skewed in beverages and other products. Related to this, Sacks et al., (2009) found no substantial effect of the introduction of a nutritional traffic light label on the sales of Ready Meals and Sandwiches. This might be a demonstration of the heterogeneity of traffic light labels between different product categories.

In a study regarding multiple binary sustainability labels, Grunert et al., (2014) found that these labels have a low level of use. The effect was not due to a low level of environmental concern as the respondents had moderately high levels of sustainability concerns. The low use level was believed to be partly caused by lack of understanding, giving reason to believe that binary labels do not provide enough information. Thøgersen & Nielsen, (2016) found something similar when investigating the carbon footprint label, a label communicating the amount of carbon dioxide equivalents (CO2e) are emitted during the life cycle of a product, on a ground coffee product. They found that the more environmentally concerned consumers were, the larger the impact of the label was. The overall modest effects of the label were amplified when they remodeled the label into a traffic light label by adding colors. In a similar research Drescher et al., (2014) not only used nutritional labelling, they included financial products. Consumers WTP for financial products increased when they were labelled with a traffic light label. Their results show that the impact of traffic light labels extends beyond the confines of food products.

In studying the effect of a 3-tiered traffic light label on seafood consumption Hallstein & Villas-Boas, (2013) found that treatment stores using this label sold significantly fewer yellow-labelled seafood, but green- and red-labelled seafood sales were unchanged. This effect was unexpected, and could have had numerous reasons, one being a correlation between types of fish and label

color. Some fish did not have a green option, for example. In their follow-up study, Hilger et al., (2019) used the same revealed preference data to measure WTP for environmental attributes of the seafood. They found that the yellow label negatively impacted WTP, needing a drop in price by a third to revert to the sales levels from prior to the introduction of the label. They hypothesize that one of the reasons for that effect is due to the over-estimation of a yellow-labelled product's environmental sustainability, when compared to green- and red-labelled products. Another study on seafood, specifically oysters, by Li et al., (2018) went in a slightly different direction. Oysters by nature can filter excess nutrients from agricultural runoff from the water. The oysters were given a gold, silver, or bronze label, depending on the agricultural runoff in the estuary where they were harvested. Oysters from estuaries with a lot of runoff where given a higher tiered label, because their filtering capabilities where cleaning the water more than oysters in waters with little to no runoff. Consumers were willing to pay more for gold labelled oysters, compared to the same oysters without label, coming from a decrease in valuation of the bronze label. These studies on seafood have differing results, which might be due to one being a field experiment in multiple supermarkets, and the other a laboratory experiment. Another contributing factor to these differing results might be that in the San-Francisco bay area, where the study of Hallstein & Villas-Boas, (2013) was held, Seafood Watch distributed pocket guides with information regarding the sustainability of certain types of fish. This potential prior information might have already influenced consumer choice before the label experiment started.

The exact nature of the effect on consumption traffic light labels is not yet properly understood, as suggested by the differing outcomes of the literature reviewed here. This study aims to contribute to the existing literature on the nature of traffic light labels by investigating the effect of introducing a tier to a two-tiered ecolabel with the same attributes, save for the label.

3. Method

Choice experiments (CE) is an important tool for analysis of environmental valuation, especially for non-use value situations (Adamowicz et al., 1998; Hanley et al., 1998, 2002). It involves the design of choice sets, which have different attributes and levels. Respondents are asked to choose between several different combinations of attributes multiple times. CE is based on random utility theory, which is a utility function that can be partitioned into two parts; one deterministic and in principle observable, and one random and unobservable (Hanley et al., 1998). The random utility model assumes that a rational individual would choose their utility maximizing alternative, within the confines of the available options.

Most of the choice experiments on traffic light labels used food products in their design. One of the problems with food-related label research is that the taste and quality of food products, and individual preferences for this quality, is very heterogeneous. More expensive foods can easily be associated with tastier, or healthier foods. Because of this potential correlation, trash bags were used in the choice design. Trash bags are literal disposables, where other than size or thickness, not many different attributes exist. A generic roll of trash bags was used, with the enviro-score label on the roll itself. See the appendix for example choice sets, including images.

The trash bags had three attributes: label, quality, and price. As stated before, the French nutri-score was used as the basis for design of the enviro-score label, using its design to create a two- and three-tiered version. In the three-tiered treatment, the left option was always the red C-label, the middle yellow B, and the right the green A-label. For the two-tiered label the left and right options were still red and green, B and A, respectively. The middle label alternated between red and green. The quality attribute had two values: Regular and Extra thick, pertaining to the thickness of the trash bags. Price, the final attribute had three levels: low, middle, and high.

Table 1
Choice experiment design: attributes and attribute levels

Attribute	Levels		
Price	Low	Middle	High
Quality	Regular	Extra Thick	
Label	Red	Yellow[1]	Green

[1]Yellow was replaced by either green or red in the three-tiered treatment group

SAS software was used to generate the choice sets for the three-tiered labelling experiment. With it I generated 32 choice sets, 4 blocks with 8 choices each, with the highest D-efficiency (Kuhfeld, 2010). Restrictions were imposed to ensure that there were no strictly dominating options, e.g. a choice where the green label was cheaper and of higher quality than the red or yellow one. Splitting the 32 choice sets into four groups of 8 was done to limit choice fatigue (Ryan et al., 2012). Afterwards, the 32 choice sets for the two-tiered experiment were made by hand, by substituting the yellow label of the three-tiered design with either a green or red label. In this design no dominating options where allowed as well. To be able to check for consistency, for 50% of the subjects the first choice was repeated after the 6^{th} set, for the other half the first choice was repeated after the 7^{th}. Each subject would thus see 9 choice sets in total. Each choice set consisting of 4 options, the fourth being "none of these" to reduce the possibility of biased estimated utility parameters (Ryan et al., 2012).

The cloud based Qualtrics software was used to create the actual surveys. Using its feature to embed data, a reference level for the price attribute was set by the subjects themselves. They were asked to select the value closest to what their household normally pays for a roll of trash bags, giving them the choice of $2, $3, $5, $10 and "I don't know". The price levels were encoded asymmetrically: level 1 is 75% of the reference price, level 2 is equal to the reference

price, and level 3 was 150% of the reference price. When “I don’t know” was selected, $5 was used as reference price.

The survey started with a couple of demographic questions, put at the start to potentially limit nonresponse (Teclaw et al., 2012). After the demographic questions, the above-mentioned price reference question was presented, followed by brief instructions. The enviro-score was explained as reflecting the alternatives’ score on all its environmental impacts related to production, transport, use, and disposal. To make sure the enviro-score was understood a control choice was added below the instruction text, where respondents had to choose the option with the highest enviro-score to be able to continue. The choice sets followed the instruction. Subjects could choose “none of these” if they would not purchase any of the options. Choosing “none of these” would result in a follow-up question, which allowed the subject to state the reason for not choosing a product. Three options were given: “Too complex to answer”, “Options were not relevant” and “Options were too expensive”. After the choice cards, subjects were asked to rank the three attributes, price, quality, and ecolabel, according to level of importance. They were then shown two products with identical quality and label, but with different prices. They were asked which of the options, if any, they expected to have the highest environmental quality. This last question is used to test the hypothesis that price is a proxy for environmental quality during information scarcity. There were four versions of this question, evenly distributed amongst respondents. Two for each color, one with regular quality for both alternatives, and one with high quality. In the appendix an example question is included of one of the four versions, as well as a simple survey flow for extra clarity.

The surveys were distributed through Amazons Mechanical Turk (MTurk) platform to residents of the United States only to make the sample at least somewhat homogeneous. Subjects were

given $0.50 upon completion, which was slightly above average for the type and length of task. MTurk gives the option to bar subjects from partaking in another of your surveys, but only after a batch is completed. Therefore, to make sure subjects could not partake in both treatment surveys, there was never more than one treatment online at the same time, and subjects were tagged afterwards to keeping them from entering again.

The first batches of participants were required to have a Mechanical Turk Masters qualification. This is a qualification Amazon hands to the top workers on MTurk, though the exact requirements are not public. 152 participants for the three-tiered treatment, and 96 for the two-tiered treatment had this Masters qualification. After this, the number of participants entering trickled to a near stop, probably because the pool of Masters is small. The Masters qualification requirement was dropped and replaced it with a 95% approval rates and a minimum of 50 completed HITs for the remaining 152 respondents for the three-tier treatment, and 208 for the two-tiered treatment.

To estimate the results, the mixed logit model was chosen, using STATA to estimate the model. Mixed logit is a flexible model to approximate random utility models. Any random utility model can be approximated by a mixed logit, if the appropriate choice of variables and mixing distribution is used (McFadden & Train, 2000). In contrast to most logit models to analyze choice experiments, mixed logit allows for random taste variation (Train, 2009). This is relevant in this experiment because the preference for quality or ecolabel is not necessarily homogeneous. Train, (2009) also clearly explains that independence from irrelevant alternatives (IIA) is not exhibited in mixed logit. IIA assumes that a change in a third available alternative does not change the relative odds of choosing between the first two alternatives. Especially in the two-

tiered model, where the one of the two label tiers are presented twice in each choice set, IIA would necessarily hold.

4. Results

Table 2 below shows the demographic composition of the respondents. The number of people in each category are grouped by treatment; treatment 1 was the group that was shown the two-tiered ecolabel, treatment 2 the three-tiered ecolabel. Because only categorical information was available, no test was performed whether the treatments differed significantly in demographics. A respondent earning $29,999 and one earning $30,000 are a whole category apart, without differing significantly. There are slightly more subjects than requested from MTurk. This is due to the nature of MTurk, if a respondent finished the survey, but did not report that back to the MTurk platform he would not get paid and would be replaced by another worker to fill the quota. Unfortunately, there was no way to track those workers down, resulting in a few unfortunate workers having done free labor. Males are slightly overrepresented, with 57% of all respondents, approximately 55% and 60% for treatment 1 and 2, respectively. As for age, 66% of the respondents in treatment 1 are between 18 and 39 years old, with 58% in the same range for treatment 2. Because of the categorical nature of this variable, no mean could be calculated for age. Approximately 70 percent of the respondents has received a bachelor's degree in college or higher, 75% for treatment 1 and 65% for treatment 2, which is significantly more than the nationwide 36% (U.S. Census Bureau, 2020). The median income for both treatments is somewhere between $50,000 and $59,999. As with age, the mean income is not known, because of the categorical nature of the variable.

Table 2
Demographics by treatment (frequency)

	Treatment 1	Treatment 2
Gender		
Male	185	170
Female	122	141
Age		
18-29	69	82
30-39	109	124
40-49	66	54
50-59	45	30
60 years or older	19	21
Income		
Less than $10,000	5	13
$10,000 - $19,999	21	19
$20,000 - $29,999	31	33
$30,000 - $39,999	41	35
$40,000 - $49,999	38	46
$50,000 - $59,999	46	57
$60,000 - $69,999	35	25
$70,000 - $79,999	23	21
$80,000 - $89,999	12	16
$90,000 - $99,999	11	14
$100,000 or more	43	24
I don't know/ want to reveal	2	8
Education		
High school graduate	72	48
Associate degree in college	38	30
Bachelor's degree in college	156	197
Graduate degree or higher	42	36
N	311	308

Table 3 shows the estimated parameters of the mixed logit models. The coefficients cannot be directly interpreted, though the relative size and sign of the coefficients does tell us some things. In the base model the treatment effects were not yet introduced, all estimated parameters are statistically significant, and preferences appear to be heterogeneous, save for price*high income. Looking at the estimates for the standard deviation of price*high income, homogeneity cannot be assumed. In model 2, price and green were interacted with treatment. In this model, again, all estimates are statistically significant, and preferences seem to be heterogeneous. Treatment was encoded as a dummy variable, 1 for treatment 2 and 0 for treatment 1. The price*treatment interaction thus shows the effect of being in treatment 2 on price sensitivity. Green* treatment therefore shows the effect of the subject being in treatment 2 on the green choice. In the third model high income was interacted with price as a control, additional on the treatment effects. High income was coded as the income categories bigger than category 5, which is $50,000-$59,999 and up. Price*high income is the effect on earning more than the approximate median of the sample on price sensitivity. Other demographic controls were tested, but not found to be significant. Ceteris paribus is implied in all interpretations below.

Table 3

mixed logit estimation results

	(1) base model	(2) w/treatment effects	(3) w/treatment effects & controls
Mean			
price	-0.808***	-0.562***	-0.789***
	(0.0556)	(0.0534)	(0.0939)
quality	0.732***	0.716***	0.706***
	(0.0953)	(0.0948)	(0.0909)
yellow	1.658***	1.852***	1.857***
	(0.113)	(0.130)	(0.121)
green	2.506***	2.050***	2.067***
	(0.135)	(0.156)	(0.158)
price*treatment		-0.402***	-0.357***
		(0.0863)	(0.0825)
green*treatment		1.253***	0.880***
		(0.227)	(0.219)
price*high income			0.284**
			(0.0977)
SD			
price	0.941***	0.883***	0.877***
	(0.0519)	(0.0535)	(0.0550)
quality	1.634***	1.577***	1.457***
	(0.111)	(0.0924)	(0.0955)
yellow	1.143***	1.624***	1.164***
	(0.136)	(0.160)	(0.119)
green	2.243***	2.102***	2.345***
	(0.142)	(0.134)	(0.132)
price*treatment		0.742***	0.554***
		(0.0575)	(0.0668)
green*treatment		1.273***	-0.800**
		(0.191)	(0.263)
price*high income			0.0643
			(0.0345)
N	14559	14559	14559

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Price

Price was encoded as the price the respondent saw, in US dollars, as explained in the previous section. The negative coefficients for price in all three models mean that respondents are price sensitive, meaning that a higher price decreases the probability of choosing an alternative. For model 2 the coefficient of price is almost half of the coefficient in the base model. This is due to the inclusion of the price*treatment variable, which splits the effect between treatments. For treatment 2 the effect of price is bigger than for treatment 1. In the third model the same holds, splitting the effect even further by including an interaction between price and high income. People with high income are less price sensitive than people with lower income. The estimate of the standard deviation suggests that this effect is homogeneous amongst the respondents.

Quality

Quality was encoded as a dummy variable, 1 being high quality, 0 being regular quality. In all models the coefficient for quality is positive. This can be interpreted as a preference for high quality, as it means that the chance that an alternative is being chosen increases when the alternative has high quality.

Yellow

Yellow is a dummy variable with red as the base category. The positive value of the coefficient in all models means that an alternative with a yellow ecolabel has a higher chance of being chosen than an alternative with the red ecolabel. This only holds for treatment 2, as treatment 1 did not contain any alternatives with a yellow ecolabel.

Green

This is the variable of most interest, together with price. Green is the highest tier of ecolabel in both treatment groups. As with yellow, it is a dummy with red as the base category. The positive value of the coefficients in all three models means that it is more likely to be chosen than red is. Or rather, the attribute green has a positive relation with being chosen, when compared to the red attribute. For models 2 and 3, green*treatment is introduced, which is an interaction between the treatment dummy and green. The positive value can be interpreted as green being more likely to be chosen in treatment 2, compared to treatment 1.

Table 4 gives the willingness to pay estimates for the attributes, for both high- and low-income levels, grouped by treatment. All willingness to pay estimates have positive signs and are statistically significant at the .1% level. The positive signs for green and yellow mean that for all four combinations of treatment and income, individuals would pay a premium for alternatives with an ecolabel with a higher enviro-score. The results suggest that low-income individuals are willing to pay $1.62 more for the yellow label than the red label. For individuals with a high income, this is increased to $2.16.

For the green label, the results are a little less straightforward. Low income respondents in treatment 1 are willing to pay $2.62 for the green label, where WTP for high income individuals from the same treatment was $4.09. For the participants in treatment 2, WTP for the green label was $2.57 and $3.42, for low- and high incomes, respectively. For low income participants in treatment 1, the WTP for green was not found to be statistically different from low income participants in treatment 2, $t(6469) = 0.1198$, $p = 0.9046$. For high income participants the same was found, no significant difference between treatments, $t(8086) = 1.0231$, $p = 0.3063$. Between low and high incomes within the same treatment it is slightly different. For both treatments, this

difference is significant at the 5% level, for treatment 1: t(7315) = -2.0532, p= 0.0401, and for treatment 2: t(7240) = -2.2252, p = 0.0261. These results indicate that it cannot be stated that there is any difference in willingness to pay for the green label when an extra tier is added to an ecolabel. Thus, no evidence is found for the hypothesis that increasing the number of tiers of a multi-tiered traffic light ecolabel will have a positive effect on willingness to pay (WTP) for the highest tiered label.

Table 4
willingness to pay estimates

	Income < $50,000		Income > $50,000	
	2-tiered label	3-tiered label	2-tiered label	3-tiered label
quality	0.895***	0.616***	1.399***	0.820***
	(0.144)	(0.085)	(0.241)	(0.104)
yellow		1.621***		2.157***
		(0.139)		(0.189)
green	2.619***	2.572***	4.095***	3.421***
	(0.314)	(0.235)	(0.606)	(0.280)
N	3357	3114	3960	4128

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To test if consumers use price as a proxy for environmental quality, all the versions of the question that asked which of two alternatives had the highest environmental quality were combined into one variable: environmental quality. B was always the most expensive option, and other than price, the alternatives where identical. None of the demographic variables were found to be significantly related to the answer to this question, using OLS. Table 5 reports the tabulation of the variable by treatment. You can see that fewer people perceived the alternatives to be equal in treatment 1, than in treatment 2. In both treatments combined, approximately 35%

of the respondents thought that the alternatives did not have the same environmental quality. There appears to be some relation between price and perceived environmental quality, which might be a mechanism that can explain differing outcomes between the two treatments. Fisher's exact of 0.00 indicates that the differences between treatments are statistically significant.

Table 5
Tabulation of environmental quality by treatment

Environmental quality	Treatment 1	Treatment 2	Total
A>B	60	27	87
A~B	178	221	399
B>A	73	58	131
Total	311	306	617

Fisher's exact = 0.00

5. Discussion

In general, the results from the choice experiment suggests that people are willing to pay extra for environmentally friendlier trash bags. The higher tier(s) were also found to be more likely to be chosen than the lowest one. There was no statistical difference in the willingness to pay for the highest tier between the treatment groups, thus rejecting the hypothesis that increasing the number of tiers of a multi-tiered traffic light ecolabel will have a positive effect on willingness to pay for the highest tiered label. Some evidence was found that price might be used as a proxy for environmental quality during information scarcity, at least for trash bags. The data for this second hypothesis was based on a single question in the survey, further study might be necessary.

Fewer respondents in the two-tiered treatment stated that the environmental quality of two products which only differed in price were the same. The lack in difference between the WTP of both treatments might be explained through the mechanism where price was used as a proxy for environmental quality. In the two-tiered treatment there was information scarcity. Because no dominating options were allowed, the average price of the green option is higher than the average price of the red option. The hypothesized effect that increasing the number of tiers of an ecolabel would lead to more consumption of the highest tier might be counteracted by the usage of price as a proxy for quality during information scarcity. Future research with treatments with a three-tiered- and a four-tiered ecolabel might find that the WTP estimates do differ significantly, because information scarcity is less present in that design.

One of the limitations of the research is inherent of stated preference, where there might be a hypothetical bias. Real life consumption behavior might differ from the hypothetical situation,

this is extremely difficult to mitigate completely. Testing a similar label to the enviro-score in an actual supermarket setting might give different results. Another limitation is that subjects were extremely aware of the label, having just had a briefing on it, and it being the largest of the attributes shown on their screen.

The sample might not be an adequate representation of the population. First, the average education level is far higher than the United States average. Second, because the requirements for being a Master on MTurk are not publicly available the consequences of relaxing this requirement are also uncertain.

Another point of interest is the limitation of comparative labelling altogether, only helping customers find the best and worst products within a product category. It might help you find the most energy efficient air-conditioning system yet does not provide any information on whether air-conditioning systems are the most efficient way to cool your house in summer. Though outside the scope of this study, combining traffic light labels with some absolute value labels, such as the carbon footprint label, might be more efficient in nudging consumers to more sustainable products (Thøgersen & Nielsen, 2016). Revealed preference research into a combination label, used on multiple different product groups would give us insight into the workings of ecolabels.

Trash bags are not a complex product and people might not have a very strong opinion about them. Though this was exactly the reason they were chosen, it still brings problems with it. The environmental impact of a roll of trash bags is of course minimal, when compared to large electronic devices. Repeating this study for products with a very high environmental impact might therefore give differing results. Especially when products within a group have a wide range in environmental impact, the addition of a tier might have a larger impact.

Though no differing effect was found between WTP for two- and three-tiered traffic light ecolabels, future research might be done to see of this would still hold for ecolabels with more tiers. The nutri-score, for example has five tiers.

6. Conclusion

The large number of ecolabels, and the ensuing lack of familiarity with these labels demonstrate the need to increase knowledge of the workings of ecolabels. This study tried to show the effect of adding an additional tier on a two-tiered ecolabel.

To investigate this effect, an online choice experiment was conducted with two treatments with hypothetical ecolabels, one with a two-tiered ecolabel, the other a three-tiered one. Using mixed logit to analyze the results, the highest tier was found to be chosen more often than the lowest one. This effect was stronger in the treatment with the three-tiered label. Willingness to pay was found to be positive and significant, though the difference between treatments was not found to be statistically significant.

Though willingness to pay was not found to differ between treatments, and my hypothesis was rejected, the study contributes a small part to a better understanding of the workings of ecolabels. Some evidence was found for the second hypothesis, that price is a proxy for quality with information scarcity, though that was limited on the data from a single question in the survey.

This effect might explain the mechanism through which consumers make their decision during information scarcity. In this study the treatment with the lowest number of tiers had only two tiers. This scarcity might influence willingness to pay enough to counteract some of the effect adding another tier had for the second treatment.

Future research on the subject could increase the minimum number of tiers of the ecolabel and use a product with a greater variety of environmental impact. Ideally using revealed preference data instead of stated preference to eliminate some of the limitations of this research. Setting up

a real-world supermarket experiment is time- and resource consuming, and therefore might not be feasible.

In conclusion, though ecolabels might be a promising tool to combat climate change, more research is needed to the workings before making changes to existing labels or adopting new ones.

References

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998). Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. ***American Journal of Agricultural Economics***, ***80***(1), 64–75.

All ecolabels | Ecolabel Index. (n.d.). Retrieved August 5, 2020, from http://www.ecolabelindex.com/ecolabels/

Brécard, D. (2014). Consumer confusion over the profusion of eco-labels: Lessons from a double differentiation model. ***Resource and Energy Economics***, ***37***, 64–84. https://doi.org/10.1016/j.reseneeco.2013.10.002

Chantal, J., & Hercberg, S. (2017). Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score. ***Public Health Panorama***, ***03***(04), 712–725.

Drescher, L. S., Roosen, J., & Marette, S. (2014). The effects of traffic light labels and involvement on consumer choices for food and financial products. ***International Journal of Consumer Studies***, ***38***(3), 217–227. https://doi.org/10.1111/ijcs.12086

Egnell, M., Ducrot, P., Touvier, M., Allès, B., Hercberg, S., Kesse-Guyot, E., & Julia, C. (2018). Objective understanding of Nutri-Score Front-Of-Package nutrition label according to individual characteristics of subjects: Comparisons with other format labels. ***PLoS ONE***, ***13***(8), 1–16. https://doi.org/10.1371/journal.pone.0202095

Grunert, K. G., Hieke, S., & Wills, J. (2014). Sustainability labels on food products: Consumer motivation, understanding and use. ***Food Policy***, ***44***, 177–189. https://doi.org/10.1016/j.foodpol.2013.12.001

Hainmueller, J., Hiscox, M. J., & Sequeira, S. (2015). Consumer demand for fair trade: Evidence from a multistore field experiment. ***Review of Economics and Statistics***, ***97***(2), 242–256. https://doi.org/10.1162/REST_a_00467

Hallstein, E., & Villas-Boas, S. B. (2013). Can household consumers save the wild fish? Lessons from a sustainable seafood advisory. ***Journal of Environmental Economics and Management***, ***66***(1), 52–71. https://doi.org/10.1016/j.jeem.2013.01.003

Hanley, N., Mourato, S., & Wright, R. E. (2002). Choice Modelling Approaches: A Superior Alternative for Environmental Valuation? ***Journal of Economic Surveys***, ***15***(3), 435–462. https://doi.org/10.1111/1467-6419.00145

Hanley, N., Wright., R. E., & Adamowicz, V. (1998). Using Choice Experiments to Value the Environment. ***Environmental and Resource Economics***, ***49***(11), 413–428.

Hilger, J., Hallstein, E., Stevens, A. W., & Villas-Boas, S. B. (2019). Measuring Willingness to Pay for Environmental Attributes in Seafood. ***Environmental and Resource Economics***, ***73***(1), 307–332. https://doi.org/10.1007/s10640-018-0264-6

Kuhfeld, W. F. (SAS software). (2010). ***SAS guide for Discrete Choice Experiments***. http://support.sas.com/techsup/technote/mr2010f.sas.

Li, T., Kecinski, M., & Messer, K. D. (2018). Behavioural responses to science-based eco-labelling: gold, silver, or bronze. ***Applied Economics***, ***50***(39), 4250–4263. https://doi.org/10.1080/00036846.2018.1441522

Ryan, M., Kolstad, J. R., Rockers, P. C., & Dolea, C. (2012). ***How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas: a user guide with case studies***.

Sacks, G., Rayner, M., & Swinburn, B. (2009). Impact of front-of-pack "traffic-light" nutrition labelling on consumer food purchases in the UK. ***Health Promotion International***, ***24***(4), 344–352. https://doi.org/10.1093/heapro/dap032

Sogn-Grundvåg, G., Larsen, T. A., & Young, J. A. (2014). Product Differentiation with Credence Attributes and Private Labels: The Case of Whitefish in UK Supermarkets. ***Journal of Agricultural Economics***, ***65***(2), 368–382. https://doi.org/10.1111/1477-9552.12047

Stadelmann, M., & Schubert, R. (2018). How Do Different Designs of Energy Labels Influence Purchases of Household Appliances? A Field Study in Switzerland. ***Ecological Economics***, ***144***(July 2017), 112–123. https://doi.org/10.1016/j.ecolecon.2017.07.031

Teclaw, R., Price, M. C., & Osatuke, K. (2012). Demographic Question Placement: Effect on Item Response Rates and Means of a Veterans Health Administration Survey. ***Journal of Business and Psychology***, ***27***(3), 281–290. https://doi.org/10.1007/s10869-011-9249-y

Teisl, M. F., Rubin, J., & Noblet, C. L. (2008). Non-dirty dancing? Interactions between eco-labels and consumers. ***Journal of Economic Psychology***, ***29***(2), 140–159. https://doi.org/10.1016/j.joep.2007.04.002

Thøgersen, J., & Nielsen, K. S. (2016). A better carbon footprint label. ***Journal of Cleaner Production***, ***125***, 86–94. https://doi.org/10.1016/j.jclepro.2016.03.098

Thorndike, A. N., Riis, J., Sonnenberg, L. M., & Levy, D. E. (2014). Traffic-light labels and choice architecture: Promoting healthy food choices. ***American Journal of Preventive Medicine***, ***46***(2), 143–149. https://doi.org/10.1016/j.amepre.2013.10.002

U.S. Census Bureau. (2020). ***Educational Attainment in the United States: 2019***. https://www.census.gov/data/tables/2019/demo/educational-attainment/cps-detailed-tables.html

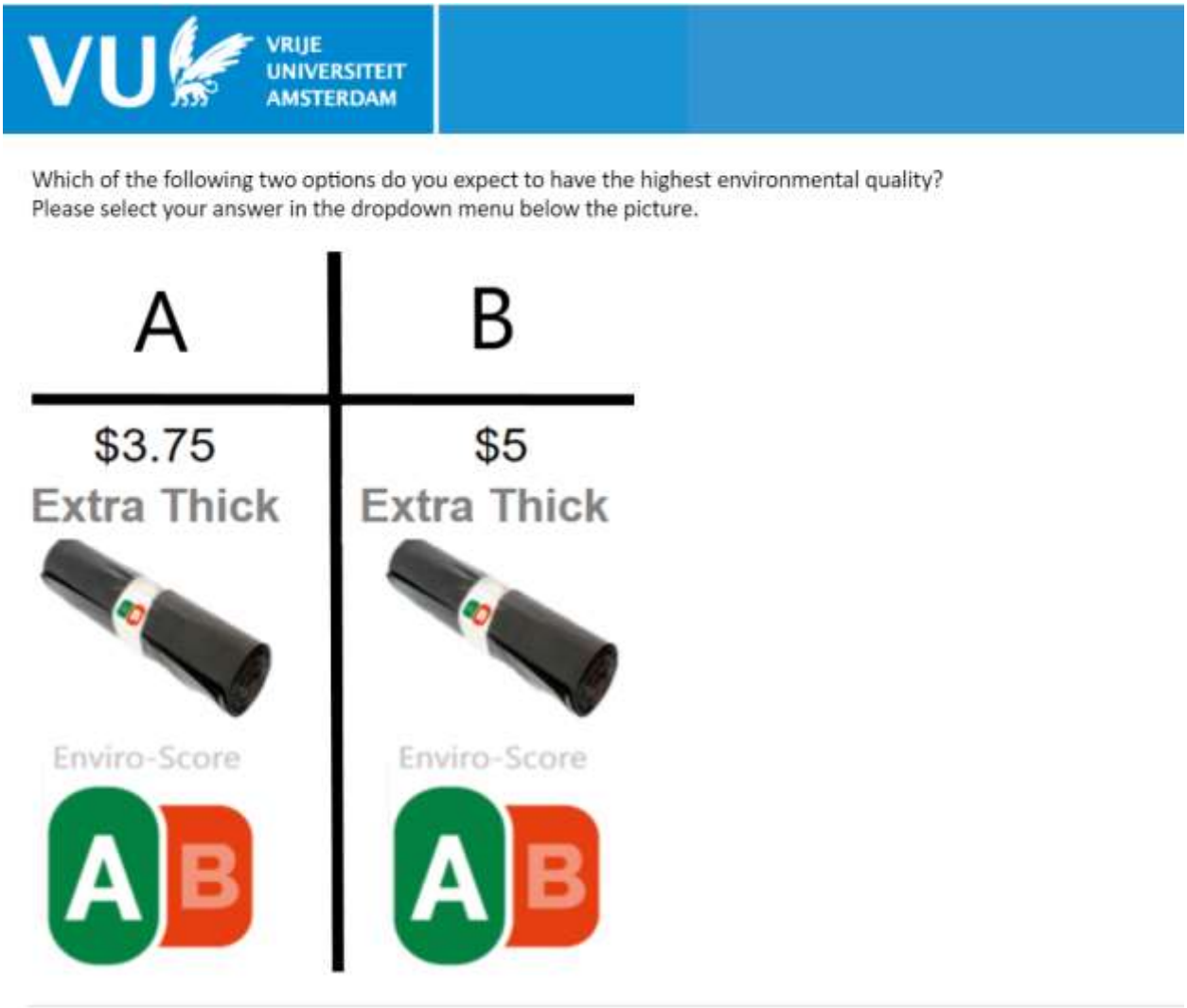
Vlaeminck, P., Jiang, T., & Vranken, L. (2014). Food labeling and eco-friendly consumption: Experimental evidence from a Belgian supermarket. ***Ecological Economics***, ***108***, 180–190. https://doi.org/10.1016/j.ecolecon.2014.10.019

Xu, P., Zeng, Y., Fong, Q., Lone, T., & Liu, Y. (2012). Chinese consumers' willingness to pay for green- and eco-labeled seafood. ***Food Control***, ***28***(1), 74–82. https://doi.org/10.1016/j.foodcont.2012.04.008

Appendix A.

Example Choice Sets

Example Price-Proxy Question



Appendix B

Survey flow