COMMISSARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE

Études & documents

n° 149
May
2016

The impact of environmental labelling on consumer choices: lessons from a large-sample choice experiment

ÉCONOMIE ET ÉVALUATION

Service de l'économie, de l'évaluation et de l'intégration du développement durable

www.developpement-durable.gouv.fr

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère de l'Environnement, de l'Énergie et de la Mer

Collection « Studies and Documents » of the Economic, Evaluation and Integration of Sustainable Development Service
(Service de l'Économie, de l'Évaluation et de l'Intégration du Développement Durable)
in the Department of the General Commissioner for Sustainable Development
(Commissariat Général au Développement Durable)

Title: The impact of environmental labelling on consumer choices: lessons from a large-sample choice experiment

Director of publication: Xavier **Bonnet**

Authors(s): Nila **Ceci-Renaud** ; Tedjani **Tarayoun**

Date of publication: May 2016

Table of contents

Executive Summary

The aim of environmental labelling is to educate consumers about the environmental impact of the products they buy. Unlike ecolabels, which are manifested by the presence of a logo on consumer products, the principle of environmental labelling is to display quantitative information on the environmental footprint of the products, in the form of a graduated scale for example. This study analyses the impact of such a scheme on consumer choices, using an original survey of 5,000 respondents with hypothetical choice experiments. We estimate consumers' willingness to pay for better environmental quality from their product choices. Our results show a high level of sensitivity to environmental quality with considerable heterogeneity of preferences between consumers. The median of willingness to pay for more environmentally friendly products would be nearly double the willingness to pay for well-known brands. Our results also show that environmental labelling is likely to reach a larger target population than ecolabels. Finally, the effectiveness of the scheme varies depending on whether it is mandatory or voluntary: when the label is not mandatory the consumer tends to consider unlabelled products as being of average environmental quality, leading producers to only reveal their environmental performance when it is above average. Environmental labelling then comes closer to ecolabels.

Introduction

Educating consumers about the environmental impact of products is an essential issue in the promotion of more sustainable and more environmentally friendly consumption. From this perspective, ecolabels[1], displayed in the form of a logo on some products to highlight their environmental excellence, have become very widespread since 1970.

Nevertheless, ecolabels have certain structural limitations in their economic efficiency: they ensure compliance with a given requirement level which is not necessarily optimum for some businesses (Spencer 1975), they do not permit comparison between ecolabelled products (OECD 2013, Brécard 2013) and the segmentation they operate in the market may lead to an increase in high quality environmental products (Zago and Pick 2004). These limitations may be exceeded by a scheme providing more complete information on the environmental quality of products. The principle of environmental labelling consists of displaying quantitative information on consumer products about their environmental quality, or a graduated scale enabling them to be situated within a product category performing the same function.

The aim of environmental labelling is twofold: firstly, consumers may transfer their purchases to products less harmful to the environment and, secondly, this re-orientation of demand encourages producers to develop more environmentally friendly methods of production. In fact, the limitations of ecolabels are partly due to the fact that there is a wide variety of virtuous practices and that the relative ease with which they can be implemented depends on each business. Economic efficiency is thus at its maximum when each business can choose where it wishes to put its efforts to reduce damage to the environment. However, it is not possible to create an ecolabel for each of these virtuous practices (which would moreover be totally unreadable for the consumer), such that most of them cannot be fully appreciated outside the company. This is even more regrettable as some practices help to reduce production costs (such as energy savings, less packaging). Indicating them could permit the emergence of a market sector of low cost green products, while ecolabels correspond to higher quality standards.

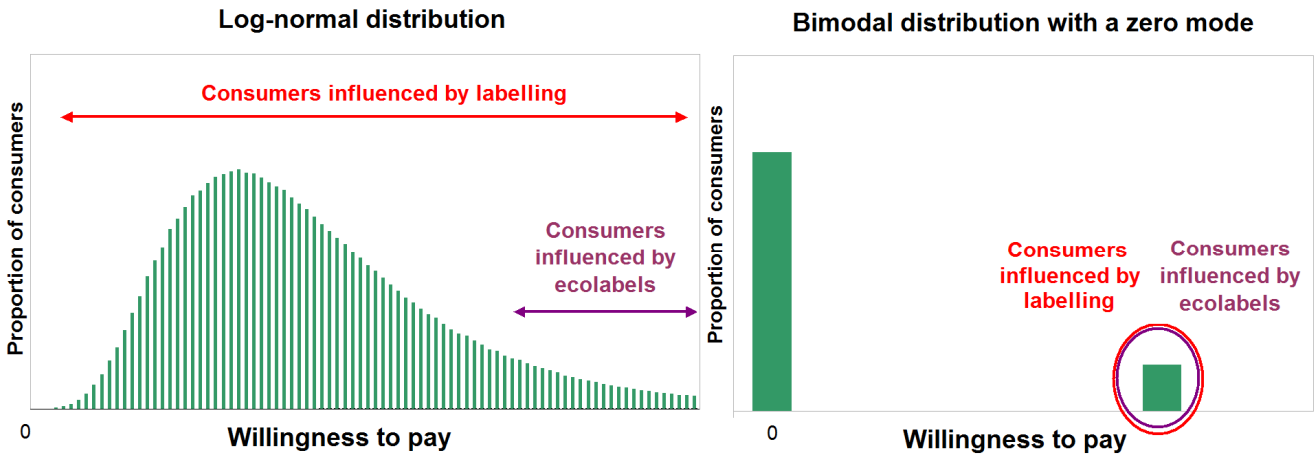
In France, the principle of a scheme for displaying environmental information on consumer products appeared in 2009 with the planning act of 3 October 2009 on the implementation of the Grenelle Environment Roundtable, article 54 of which states that "*consumers must have access to sincere, objective and comprehensive information on the overall characteristics of the products and packaging*» for ordinary consumer products. Thus, in the spirit of this article, environmental labelling is not limited to the carbon footprint alone but must include relevant environmental impacts for a given product category (pollution and water consumption, exhaustion of resources, loss of biodiversity, etc.), calculated over the whole of their life cycle. Against this background a national experiment was carried out in 2011 and 2012, in the course of which 168 volunteer businesses set up their own labelling schemes in line with the above principles. It was the subject of a report to the French parliament[2]. These aims also inspired the European Commission, which is carrying out its own experiments on the issue, with a view to the possible general implementation of an environmental labelling scheme at the European Union scale.

In view of the fact that numerous environmental logos already exist, the impact of an environmental labelling scheme on consumer behavior depends on the option of implementation. A purely voluntary scheme would above all serve as a "comparator" between labelled products, while mandatory labelling would enable not only the "best" products to be identified but also the "worst" and would be likely to lead to more consumer reports (see box 1). In this latter case however, the ability of environmental labelling to have more influence on consumer behavior than ecolabels depends on the distribution of willingness to pay (WTP) for the environment among consumers. This may be illustrated by two extreme cases (figure 1): if there are only two categories of consumers, with a majority of consumers totally insensitive to the environment, the target of environmental labelling is the same as that of ecolabels; on the other hand, if the distribution of willingness to pay is diffuse, the target of environmental labelling is much larger than that of ecolabels. The latter currently only exploit the highest willingness to pay for the environment by indicating the environmental excellence of their products. Environmental labelling on the other hand would enable consumers with lower willingness to pay to express their preferences by choosing an intermediate environmental quality.

1 i.e. type 1 environmental declarations according to ISO terminology.

2 http://www.developpement-durable.gouv.fr/Bilan-au-Parlement-de-l.html

Figure 1 : Target of a compulsory labelling scheme depending on the distribution of willingness to pay



Due to the unusual nature of the environmental labelling scheme, there are few empirical results regarding its impact on consumer behavior. Vanclay et al. (2010) in-store experiment on the implementation of a "*traffic light*" system just for the carbon footprint and observing significant changes in market share over a limited range of products. Bertrandias et al. (2012) show moreover that the strength of consumer reaction to environmental labelling varies according to their sensitivity to price and their perception of the environmental issues of the act of purchase.

The average level and the distribution of willingness to pay for environmental quality of products are also poorly known owing to the difficulty of measuring them[3]. Surveys that include direct statements of willingness to pay are deemed to include a significant declarative bias; laboratory experiments with real payments mainly concern limited samples, unrepresentative of the general population; finally in-store experiments enable to measure impacts on market share but rarely willingness to pay, much less its distribution.

To supplement this work, the French Ministry for Sustainable Development designed and funded a survey using hypothetical choice experiments to assess consumer reaction to environmental labelling on three families of consumer product.

3 For more details on existing methods for measuring willingness to pay see box 3 of Ceci-Renaud and Thao-Khamsing (2012),"Les consommateurs face à l'affichage environnemental", Études et documents n°74, November 2012, Commissariat général au développement durable.

Box 1: French national experiment on environmental labelling and implementation options

The national experiment on environmental labelling carried out in 2011 and 2012 involved 168 volunteer businesses with various labelling formats (figure E1). The experiment report presented to the French Parliament identifies three options for general implementation of the scheme, which are likely to have different impacts on consumer behavior. Regardless of the option chosen, the labelling scheme would have a standardized format in terms of content and visual appearance and would be designed as a supplement to the pre-existing environmental logos. It is the more or less restrictive nature of the scheme that differs between the three implementation options:

- A voluntary labelling scheme would leave businesses entirely free to carry out labelling. In this case, the scheme would resemble a comparison tool for labelled products.
- A restricted voluntary labelling scheme would impose environmental labelling on businesses once they announce the environmental performance of their products. The scheme would then constitute a tool for combating the abusive use of green marketing ("*greenwashing*") ;
- A mandatory labelling scheme would be imposed on all products of certain predetermined categories. In this case labelling would enable the best and worst environmental performances to be identified and would be a consumption guidance tool.

Figure E1 : Examples of environmental labelling formats used in French experiments in 2011-2012

1. Design and completion of the choice experiments survey

The survey studied the way in which consumers assess the environmental characteristics of different consumer products, according to different forms of labelling (mandatory, voluntary, etc.). Conducted over the *internet*, the survey places the respondent in the context of an on-line purchase, in which boxes of washing powder, yoghurt and jeans are offered for sale, with each product bearing a label summarizing its environmental impact (over its life-cycle). The sample, selected in November 2013 according to the quota method, consisted of 5,246 representative members of the French population aged 18 and over[4].

In each experiment, known as a "choice experiment", the respondent is faced with three products of the same type - washing powder, yoghurt or jeans -and selects the one he would prefer to buy (figure 2). The experiment is carried out three times for each product type, i.e. a total of nine choice experiments per respondent.

The products on offer have different characteristics, randomly selected according to uniform distributions:

- the product brand name, which is either a high commercial capital value[5], or a fictitious brand ;
- The product price, between a credible minimum and maximum[6], which is set by the respondent at the start of the questionnaire
- The volume, for the washing powder only, which is between 2 and 3 liters

[4] Researchers may request data by email from: Ernr.Seei.Cgdd@developpement-durable.gouv.fr. Further description is available in the French version of this paper.

5 The brand names were deliberately removed from each publication, in the interests of neutrality.

[6] The interval selected by the respondent falls within absolute boundaries: from 1 to 16 euros for the four yoghurts, 2 to 25 euros for the washing powder, 10 to 500 euros for the jeans.

- An environmental label, combining three environmental criteria: "CO2 emissions" which assesses the product's effects on global warming and a selection of two other environmental criteria relevant to the product category concerned7. Each criterion takes a (fictitious) value between "A" and "E".

The visuals (product shape, label format, colors, etc.) remain unchanged between products, in order to limit the number of characteristics to be controlled in the econometric models and the resultant loss of precision.

The selection of prices within a range set by the respondent make the purchase more realistic in relation to the consumer's normal budget for the products. This is particularly important since the consumers are asked to imagine that the non-random characteristics of products are identical to those of their usual products. This corresponds to a virtual situation in which each product range includes several references that might differ in their environmental footprint. In practice, the intervals chosen by the respondents are quite wide, which provides sufficient price variability to enable a good identification of willingness to pay. For each product family, fewer than 5% of respondents have a difference of less than 20% between their minimum and maximum prices, and 5% of respondents have a difference greater than a factor of 5. The median is between a factor of 1.6 and a factor of 2 depending on the product family.

The product attributes are selected independently, except for the three environmental criteria, which are imposed with a correlation of 80%. In fact, in the absence of correlation, the average of the three criteria would be much less variable than the criteria taken in isolation, such that the overall ratings "E" and "A" would be less numerous (less than 7% of products). Moreover, eco-design processes generally tend to improve several environmental dimensions at the same time. The work of Ruffieux et al. (2014) carried out from a food products catalogue showed correlations of this order of magnitude between global warming footprints, water pollution and air acidification.

Figure 2 : Example of choice experiments in the reference version

[7] Thus, for washing powders, the chosen criteria, apart from CO_2 emissions, are water pollution (ecotoxicity)and water consumption. For the jeans, they are CO_2 emissions, water pollution (eutrophication) and renewable energy consumption. For yoghurt, they are CO_2 emissions, water pollution (eutrophication) and loss of biodiversity (intensification of dairy farming). These criteria illustrate the environmental pressures exerted by these products over their life cycle, in the production phase as well as in the use or disposal phase. Cf. appendix 8 for details on the criteria.

Four versions of the questionnaire were prepared, to test four types of implementation of environmental labelling. Respondents were randomly given one of the four versions:

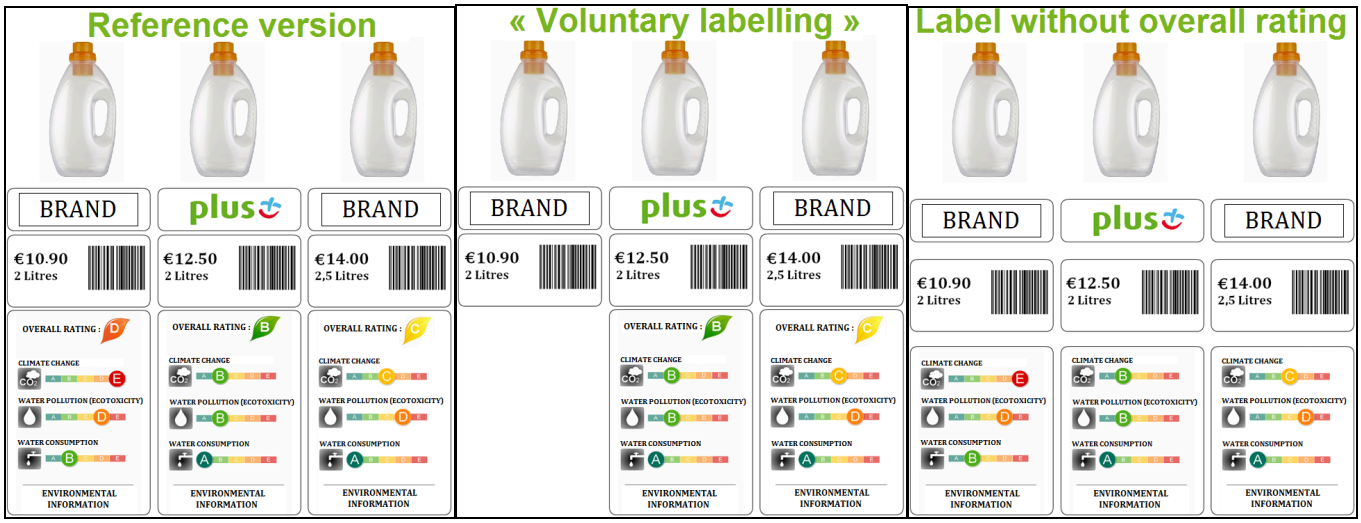
(1) reference version: each product has an environmental label accompanied by an overall rating which summarizes the product's performance. It corresponds to the simple average of the ratings awarded to the three environmental criteria.

(2) voluntary labelling version: products displaying their environmental label are only those with an overall rating from "A" to "C"

(3) version without overall rating: each product has an environmental label but the overall rating is not displayed

(4) version with educational material: similar to the reference version but an information page about the environmental impacts of the products precedes the choice experiments (see appendix 4).

Figure 3 : Example of choice experiments in the different versions of the questionnaire



A list of questions was asked at the end of the choice experiments in all versions of the questionnaire. They tell us about the respondents' opinion of the current environmental information as they come across it in their daily lives (clarity, usefulness, etc.), their habits in terms of sustainable consumption and their awareness of ecological issues. These questions are asked after the choice experiments so as not to influence the product choices. In total, the average time for answering the questionnaire is 20 minutes.

2. Econometric model and estimation

As explained in the introduction, the issue of this paper is to find out not only the average, but also more broadly, the distribution of willingness to pay for the environmental quality of products among consumers. This is why our modelling gives substantial leeway to individual heterogeneity without in anyway being interested in the factors determining individual preferences. These factors could be the subject of later work due to the numerous questions in the questionnaire that could throw light on this subject. It is even more important to consider individual heterogeneity as it is very significant and ignoring it could bias the estimation results in a non-linear model. However, this richness of the model created estimation problems that we resolved by using a hierarchical Bayesian approach.

2.1 Model specification

In a perfect rationality model, an individual offered a discrete choice (i.e. relating to a finite number of mutually exclusive actions) has the capacity to rank all possible choices in order of preference, and necessarily chooses the one that maximizes his preferences. In practice, an individual's behavior may deviate from perfect rationality, his choice appearing to be subject to a random mechanism. This type of behavior is comprehended by the so-called discrete choice models, grouped into two separate large families. The first supposes that the decision rules are determinist and utilities random (random utility models) while for the second, the decision rules are random and the utilities determinist (random decision-rule models).

In random utility models the random nature of the utility may give expression to different realities. On the one hand, it may reflect the modeller's inability to precisely comprehend the causes that determine the choice (omitted variables). In this case, perfect rationality is not necessarily broken by the individual but the modeller does not manage to reproduce it correctly. On the other hand, the individual's utility may be influenced by internal or external parameters, forming his "state of mind". These

two interpretations may however combine in breaking down the individual's utility into a collection of observable characteristics and a random term, reflecting both the imperfect information of the modeller and the influence of the individual's state of mind.

In random decision-rule models people are not always sure of the alternative they should select and do not always make the same choice under apparently identical conditions (Tversky 1972). In fact, the reality of choices is complex and in practice it often calls for a simplification of the decision-making process. For example, when individuals are confronted by too many attributes or too many alternatives they can simplify their reasoning by hiding certain attributes or even some alternatives. This behavior is finely modelled in random decision-rule models. One of them, the "elimination-by-aspects" model proposed by Tversky (1972), assumes that the choice of an alternative is the result of a random process of elimination of the other alternatives according to the presence or absence of certain characteristics. At the end of the process, when no more elimination is possible, the remaining alternatives have the same probability of being chosen. In the same way, in order to provide an adequate description of reality, some authors have carried out empirical studies of semi-compensatory models[8], in which individuals start by limiting the range of possibilities on the basis of thresholds on certain dominant attributes, then actually compare all the attributes for the remaining alternatives (Cantillo and Otuzar 2005, Campbell 2014).

In our case, the complexity of the decision process is limited, as individuals have to choose one everyday consumer product out of just three, and the products only differ by a maximum of four attributes, with a limited number of details. In addition, individuals are only offered products belonging to a price range they have specified beforehand, such that immediate elimination behavior relating to the price criteria is unlikely. Consequently, and for a more direct analysis of the distribution of willingness to pay, we prefer to use a random utility model, i.e. a purely compensatory model in which products are systematically assessed across all their characteristics.

Following the work of McFadden and Train (2000), we specify a mixed logit model which allows a random variation of utility parameters from one individual to another. Since we are interested in willingness to pay distributions, we use utility adjustment in the willingness to pay space rather than the preferences space.

The model is written:

$$U_{itj} = \alpha_{i1} 1_{j=1} + \alpha_{i2} 1_{j=2} + \lambda_i \left[WTP_{rating,i}\, RATING_{itj} + WTP_{Washing\ powder\_\ Brand,i}\, 1_{Washing\ powder\_\ Brand,\ itj} + WTP_{Jeans\_Brand,i}\, 1_{Jeans\_Brand,itj} \right.$$

$$\left. + WTP_{YoghurtBrand,i}\, 1_{Yoghurt\_Brand,itj} + WTP_{volume,i} \log(\text{Volume}_{itj}) - \log(p_{itj}) \right] + \varepsilon_{itj} \qquad (1)$$

in which each individual i in the course of the choice experiment t chooses the product j (j = 1 to 3 according to its positioning to the left, center or right) which maximizes its utility U_{itj}. The characteristics influencing the choice are the following:

- The lateral positioning of the product: j=1 or j=2 if the product j is situated to the left or center respectively
- The non-monetary attributes of the product: the product's overall rating, linearly included in the model (A=5, B=4, C=3, D=2, E=1), one of three famous brands (depending on the product family) and the logarithm of the volume of the product (variable for the washing powder only, implicitly equal for the other products)
- The product price p_{itj} (taken in logarithm).

The environmental rating is introduced linearly, such that the associated parameter is interpreted as willingness to pay for a one-step increase in environmental rating (such as E to D or D to C). In fact, additional estimations show that consumers tend to understand the overall rating linearly: willingness to pay for a grade A rather than B is very close to willingness to pay for a grade B rather than C, etc. (see appendix 1).

Willingness to pay for the brand is differentiated by product family, which allows different preferences for each brand. Conversely, willingness to pay for the environmental rating of the product is presumed to be independent of the product family. It expresses an appreciation of environmental quality *per se*, not conditional upon the product concerned. These hypotheses were tested right.

When the other parameters remain unchanged (i.e. constant willingness to pay), parameter λ_i measures the intensity with which individuals' choices are sensitive to all characteristics of the product (apart from lateral positioning). In particular, it may reflect more or less care given to the choices made. Finally, the error term ε_{ijt} is assumed to be identically and independently distributed (i.i.d), according to a Gumbel Distribution $\pi^2 / 6$.

[8] A model is said to be compensatory when the buyer compensates for bad scores on some characteristics of the product with good scores on other characteristics of the product.

In the particular case in which the sample studied is subject to the "voluntary labelling" version, we introduce a parameter δ_i for the willingness to pay for unlabelled products:

$$U_{itj} = \alpha_{i1} 1_{j=1} + \alpha_{i2} 1_{j=2} + \lambda_i \left[WTP_{rating,i} \, Rating_{itj} + \delta_i 1_{\{\text{unlabelled product}\}} + WTP_{Washing\ powder\_Brand,i} 1_{Washing\ powder\_Brand,itj} \right.$$

$$\left. + WTP_{Jeans\_Brand,i} 1_{Jeans\_Brand,itj} + WTP_{YoghurtBrand,i} 1_{Yoghurt\_Brand,itj} + WTP_{volume,i} \log(Volume_{itj}) - \log(p_{itj}) \right] + \varepsilon_{itj} \quad (2)$$

The ratio of δ_i to the willingness to pay for the environmental rating reveals the environmental rating that the individuals implicitly attribute to unlabelled products

To complete the specification of the model, we must chose an appropriate distribution for each of the random coefficients. The normal distribution is not very suitable for willingness to pay, which cannot be negative (unless we think that some individuals want environmental damage). This is why we test other types of distribution, namely a normal distribution censored below zero, a log-normal distribution and a Johnson S_B distribution. These three distributions, which are obtained by transforming an underlying normal distribution, are characterized by uniquely positive or null values. For each individual parameter, we select the distribution that maximizes the likelihood of the model.

2.2 Method of estimation

The models are estimated in a hierarchical Bayesian method (cf. box 4). One of its advantages is that it does not require maximization of the likelihood function and thus avoids numerical difficulties. Another way of getting round these difficulties would be to estimate a simulated maximum likelihood. Anyway, Scarpa, Thiene and Train (2008) show that these two methods of estimation lead to very similar results, for models formulated in both the preferences space and the willingness to pay space.

As the economic literature gives little information on the distribution of the different willingness to pay, we select very diffuse prior distributions for the meta-parameters (i.e. uninformative prior distributions). Even with less diffuse prior distributions, the large size of the sample renders the results qualitatively robust to the means of prior distributions. The Bayesian estimator, calculated as the mean of the posterior distribution of the parameters, is asymptotically equivalent to the maximum likelihood estimator. Its variation can be estimated by the empirical variance of the posterior distribution. Consequently, the results obtained by Bayesian inference may be interpreted according to a frequentist model. This is what we do in the rest of this document, by giving the estimated value and the standard error for each estimated parameter.

Our estimates take into account the sample weighting, representative of the French population over 18 years of age[9]. The results are not significantly different with a weighting representative of product purchasing, each respondent's weighting being multiplied according to purchasing frequency.

[9] It is difficult to take into account the sample weightings in simulations. To retain the representativeness of the weighted sample in our estimates, we work from a duplicated sample: we create as many identical individuals as their sample weighting multiplied by 10. The accuracy of estimates is then corrected by this same factor of 10.

Box 2: Hierarchical Bayesian estimation method for logistic models with preference heterogeneity (mixed logit)

This model is described in detail in Train's work (2003), chapter 12.

The Bayesian approach consists of representing the uncertainty on the parameters of the model by probability distributions on them. *Prior* distributions are postulated according to what the literature, theoretical modelling or even good sense inform us about the parameters. When the data provides information about the parameters of the model, the distribution of parameters conditional upon observations, known as *posterior* distribution, is different to the *prior* distribution.

For a quadratic loss function, the Bayesian estimator minimizing the mean of the loss is simply the mean of the *posterior* distribution of the parameters. From a frequentist perspective, this estimator has asymptotic properties similar to the maximum likelihood estimator, and its variance can be estimated by that of the *posterior* distribution of the parameters.

The "hierarchical" estimation method consists of treating each individual effect of the model as a parameter to be estimated. There are therefore individual and global parameters (i.e. a kind of hierarchy between parameters). Global parameters include the homogeneous parameters of the model (those that are identical for all individuals) and hyper-parameters which characterize distributions of individual effects. In our model, hyper-parameters are the mean and the variance of the vector of normal distributions underlying the individual effects. There are no homogeneous parameters in our model.

The hierarchical Bayesian estimator of meta-parameters is calculated from an arbitrarily large number of samples from the *posterior* distribution of parameters. It is then sufficient to calculate the empirical average and variance of these samples to obtain the Bayesian estimators of any function of the parameters as well as its variance.

A Gibbs process is used to simulate the joint distribution of individual and meta-parameters: we alternately sample from the distribution of individual parameters conditioned on global parameters and on the data and from the distribution of global parameters conditioned on individual parameters. The process converges on samples from the joint *posterior* distribution of parameters for any initial value chosen on the density support. In practice, we must also use the Gibbs process to sample the global parameters, with separation in two stages for average and variance meta-parameters. The sampling of individual parameters requires the use of the Metropolis-Hastings algorithm, with a target acceptance rate at 0.3.

We use very diffuse prior distributions for these simulations, in which dispersion is so large that means do not exist. The meta-parameters of the mean follow a normal distribution centered on zero with an infinite variance (improper distribution). The meta-parameters of variance follow an inverse-Wishart distribution (distribution supported by positive-definite symmetric matrices) with an identity matrix and as many degrees of freedom as random parameters (hence mean is undefined).

3. Results

We start by giving a detailed presentation of the results obtained from the reference version of the questionnaire, then we compare the different versions to draw conclusions on the labelling option most likely to influence consumer behavior. Finally, we see how willingness to pay for the environmental performance of products is linked to ecolabel purchase behavior.

3.1. High degree of sensitivity to the environmental quality of products among respondents

As stated earlier, we tried in turn several distributions for the random coefficients (normal distribution, censored normal, log-normal or Johnson SB). The quality of adjustment of the model is assessed by calculating the corresponding log-likelihood10. This log-likelihood, while not being used in Bayesian inference, provides a frequentist choice criterion between models.The most suitable distributions are the following:

- A normal distribution for lateral preferences α_1 and α_2;
- A censored normal distribution for willingness to pay for the volume of washing powder (WTP_{volume});
- A log-normal distribution for the other individual parameters[11]: λ, WTP_{rating}, $WTP_{Washing\ powder\_Brand}$, WTP_{Jeans_Brand} and $WTP_{Yoghurt\_Brand}$.

Table 1 shows the estimation results of the accepted model (log-likelihood = -7,379) for respondents in the reference version of the questionnaire.

[10] Log-likelihood is simulated with 1,000,000 samples and calculated to the estimated value of underlying normal distribution parameters.

[11] In estimating the various models, we assumed for the sake of consistency that the willingness to pay parameters for the brand ($WTP_{washing\ powder\_brand}$, WTP_{jeans_brand} and $WTP_{yoghurt\_brand}$) fell within the same type of distribution.

First of all we find that respondents have a high degree of sensitivity to the environmental quality of products: median willingness to pay[12] for a higher overall rating is 20% of the product price. It exceeds the median willingness to pay for well-known brands (6% to 11% of the price depending on the product). Willingness to pay related to the volume of the product is lower than what should prevail in a rational model: for the median consumer, a 10% increase in the volume of washing powder is manifested by just a 5% increase in willingness to pay. We should also point out that the visuals of the washing power pack were the same size for the 2L and 3L bottles, potentially causing a downward bias in the willingness to pay for volume.

Moreover, a majority of consumers have a slight preference for products positioned to the left or center in the choice experiments, which is related to the direction of reading: respondents generally start by considering the product situated furthest to the left. The influence of the lateral positioning of products may also reflect the decision-making process in which the respondent selects the first product that comes within a certain number of predefined criteria, rather than maximize utility by considering all three products. It is a fairly marginal phenomenon, which is insufficient to cast doubt on the choice of a random utility model. In fact, for the median consumer, preference for the product on the left relative to the product furthest to the right is equivalent to a price differential of 3.3 %[13].

In addition to analysis of the median willingness to pay, the distribution of willingness to pay, both for environmental quality and for well-known brands, reveals a high degree of heterogeneity among respondents: especially in relation to environmental quality, one quarter of respondents have a willingness to pay that is lower than 6% of the product price, while for another quarter it is more than 68 % of its price[14].

Willingness to pay for the environmental rating, the brand name and the volume have a strong positive correlation, indicating that individuals sensitive to the environmental quality of products are also sensitive to well-known brands and to volume (cf. Appendix 2). Parameter λ_i, which measures the impact of product characteristics on decisions, is negatively correlated to every willingness to pay parameter: individuals who take the most care in their choices are also those who above all seek the minimum price.

Table 1 : Estimation of quartiles for each parameter of the model (standard deviation in brackets) in the reference version of the questionnaire

Parameter	Type of distribution	Average	Standard deviation	Quartiles		
				1st quartile	Median	3rd quartile
α_1 *left (ref. right)*	Normal	0.22 (0.06)	0.76 (0.07)	-0.29 (0.07)	0.22 (0.06)	0.7367 (0.07)
α_2 *centre (ref. right)*	Normal	0.20 (0.04)	0.35 (0.05)	-0.03 (0.06)	0.20 (0.04)	0.44 (0.05)
λ	Log-normal	17.5 (1.5)	42.4 (6.6)	2.6 (0.2)	6.7 (0.4)	17.0 (1.1)
WTP_{rating}	**Log-normal**	**100 % (14 %)**	**490 % (128 %)**	**6.0 % (0.5 %)**	**20.2 % (1.3 %)**	**67.6 % (5.3 %)**
$WTP_{Washing\ powder\_Brand}$	Log-normal	214 % (86 %)	7,251 % (6569 %)	1.1 % (0.3 %)	6.5 % (1.2 %)	38.5 % (5.4 %)
WTP_{Jeans_Brand}	Log-normal	217 % (71 %)	6,281 % (4,371 %)	1.3 % (0.3 %)	7.6 % (1.1 %)	43.7 % (5.4 %)
$WTP_{Yoghurt\_Brand}$	Log-normal	110 % (30 %)	1,116 % (697 %)	2.6 % (0.6 %)	10.9 % (1.8 %)	46.4 % (5.7 %)
$WTP_{washing\ powder\ volume}$	Censored normal	78.7 % (5.7 %)	89.2 % (7.0 %)	0	9.5 % (10.2 %)	134 % (9.2 %)
			Log-likelihood = -7,379			

Interpretation: willingness to pay for the environmental quality of the product is log-normally distributed in the population, with a median value equal to 20.2% of the product price.

Note: the parameters are estimated by hierarchical Bayesian estimation, according to normal, log-normal or censored normal distributions (column 2 of the table). The resulting log-likelihood is estimated to be - 7,379. The types of distribution chosen for the parameters are chosen to maximize it.

[12] To assess the different willingness to pay we have chosen to comment on the medians and quanties rather than averages. In fact, the log-normal distributions followed by some random coefficients have high extreme values (right heavy tailed distribution) which strongly influence the average value but concern few individuals. In addition, due to the nature of the distributions chosen, the medians of coefficients correspond to the average values of the underlying distributions: the median consumers from the point of view of interest coefficients is the average consumer from the point of view of underlying normal distributions.

[13] 0.22/6.7

[14] Later work will set out to identify a typology of attitudes to the environment within this willingness-to-pay heterogeneity.

3.2. Willingness to pay for environment is sensitive to the labelling format

Four versions of the questionnaire were created, according to different forms of labelling, and each respondent has been given one of the four versions. This enables us to compare willingness to pay obtained on sampling subject to the reference version of the questionnaire (part 4.1 above) with that obtained from samples subject to the other versions.

Mandatory labelling would be more effective than voluntary labelling in terms of impact on consumer buying behavior

Our study shows that mandatory labelling would be more effective than voluntary labelling in terms of impact on consumer buying behavior. Firstly, environmental labelling has more influence on consumer choices in a mandatory framework than in a voluntary one. Within products displaying a label, willingness to pay for environmental quality is higher in a mandatory framework than in a voluntary one (20% and 13% of the product price respectively, table 2). This result can have several interpretations. The presence of negative signals (below average ratings) in a mandatory framework may for example lend credibility to the environmental information provided by the labelling scheme. These negative signals, delivered on a red background, may also increase consumer perception of the product's potentially harmful effect on the environment.

Secondly, in a voluntary labelling framework, consumers consider products without environmental labels to be only slightly inferior to products displaying a label. In fact, the median consumer implicitly gives unlabelled products an overall rating equal to 2.5 [15] (standard deviation equal to 0.04), i.e. equivalent to a rating between C and D. This implicit rating is just below the displayed range of ratings (A to C) while the rating scale goes down to the letter E. This result is a *posterior* validation of our hypothesis according to which only ratings above or equal to the average would be displayed in a voluntary framework: a grade D or E reduces the probability of a product being purchased by the median consumer, such that it would probably not be displayed. In a voluntary framework environmental labelling would therefore not spread spontaneously across products with below-average environmental quality.

It is possible that the implicit rating given to unlabelled products directly depends on ratings appearing on labelled products. In fact, consumers were confronted with rating scales on domestic appliances going lower than the range of products actually offered for sale. If the consumer only sees high ratings, he might think that the majority of products are situated in the displayed range. In this case, multiple equilibrium may exist in a voluntary framework. For example, if labels are displayed only on products labelled A or B (instead of A, B and C), the implicit rating consumers give to unlabelled products may go up, such that it would no longer be beneficial for C-rated products to display a label. In such a situation, labelling may never spread spontaneously: if the first products to display their label are A-rated products, it will not be in the interest of other products to display labels. The labelling scheme would then be no different from type I ecolabels.

As expected, education about the environmental impacts of products has a positive effect on willingness to pay for a higher environmental rating. The median willingness to pay is 25% of the product price after education, compared with 20% without it (reference version). This difference is relatively small, but significant at the threshold of 5%. It should be noted that the education considered here corresponds to second level education, which describes the specific environmental impacts of the yoghurt, washing powder and jeans presented in the choice experiments (see appendix 4). Environmental labelling is explained to respondents in all versions of the questionnaire. This constitutes first level education (see appendix 3).

When environmental labelling has no overall rating, the median willingness to pay for environmental quality is slightly lower[16] (18 %, compared with 20 % in the reference version with overall ratings). The difference is however only significant at the threshold of 10%. In the context of our survey, an overall rating summarizing environmental information to rank products more easily proves of little use. The experimental conditions directly influence this result: environmental footprints displayed on the products are 80% correlated. Therefore, situations in which the consumer really needs an overall rating to determine the best product (or the worst) are not very frequent: only 20% of choice experiments. In the absence of correlation this rate would be 75%. The scope of our result under real conditions is therefore dependent on the actual correlation between the different environmental footprints of the products.

[15] This implicit rating is obtained by comparing the median willingness to pay for unlabelled products with the median willingness to pay for the overall rating (δ_i /$WTP_{rating,i}$). The median consumer considered here is median for δ_i and $WTP_{rating,i}$ and not for the ratio.

[16] In the "without overall rating" version, the overall rating is calculated and introduced into the model (it is the simple average of the various criteria), but it is not displayed on the product.

Table 2 : Estimation of WTP for the environment according to the type of labelling (standard deviation in brackets)

Version of the questionnaire	1st quartile of $CAP_{note,\ i}$	Median of $CAP_{note,\ i}$	3rdquartile of $CAP_{note,\ i}$
Reference (overall rating, no education, mandatory labelling)	6.0 % (0.5 %)	20.2 % (1.3 %)	67.6 % (5.3 %)
Display without overall rating	5.1 % (0.4 %)	17.6 % (1.1 %)	60.5 % (4.6 %)
With education	7.3 % (0.5 %)	25.1 % (1.5 %)	86.2 % (6.9 %)
Voluntary labelling	3.7 % (0.3 %)	12.2 % (0.8 %)	40.8 % (2.9 %)

Interpretation: willingness to pay for the environmental quality of the product has a median value of 25.1% of the product price in the experiment with education.

Willingness to pay for environment appears similar with regard to the product family

Willingness to pay for a higher environmental quality does not significantly vary according to the product family considered (yoghurt, washing powder or jeans) (table 3). In our choice experiments willingness to pay for the environment is proportional to the product price, regardless of whether it is a relatively occasional product (jeans), an everyday purchase (yoghurt) or known for its environmental toxicity (washing powder). However, this result could be influenced by the fact that our survey does not reproduce the more or less routine nature of purchases depending on the product (cf. Part 5).

Analysis by product confirms that median willingness to pay for a higher overall rating exceeds median willingness to pay for any of the brands with high commercial value used in the survey. This result tends to confirm that there is a margin of action for deflecting consumption and production towards more sustainable practices: budgets currently devoted to eco-design of products are probably less than those for advertising, while a suitable labelling scheme would make these investments just as profitable.

Table 3 : Estimation of median WTP according to product family (standard deviation in parenthesis)

Product	WTP Rating	WTP Brand	WTP Volume
Yoghurt	19.2 % (1.4 %)	9.2 % (2.1 %)	-
Jeans	20.2 % (1.9 %)	9.7 % (2.7 %)	-
Washing powder	18.9 % (1.6 %)	4.8 % (1.0 %)	34.6 % (13.8 %)

3.3. Environmental labelling, more effective than ecolabels

The willingness to pay for environmental quality is strongly related to the purchase of eco-labelled products: the median WTP changes from 6% of the product price for consumers who never buy eco-labelled products to 47% for consumers who often buy them (table 4). However, this WTP is very significant even for respondents who never buy eco-labelled products (18% of the population). Even supposing an upward bias in willingness to pay revealed in our experiments (see part 5), these results tend to show that the majority of consumers have a positive willingness to pay for more environmentally friendly products. Environmental labelling would therefore tend to reach a broader spectrum of consumers than the limited markets of ecolabels.

Table 4 : WTP for the environment according to the consumption of eco-labelled products

Frequency of purchase of eco-labelled products	Distribution in the population	Median of WTP / IC 95%
Always or often	23 %	46.8 % [36.4 %; 57.2 %]
Sometimes	51 %	20.2 % [17.0 %; 23.3 %]
Never	18 %	6.4 % [4.2 %; 8.5 %]
Don't know/ Don't do the shopping	8 %	20.4 % [12.0 %; 28.8 %]

4. Discussion: measurement bias and scope of results under real conditions

Carrying out a survey with hypothetical choice experiments enables us, as we have just seen, to estimate consumers' willingness to pay for various product characteristics. The results obtained are however subject to a hypothetical bias due to the virtual nature of the choice experiments carried out during the survey (cf. Ceci-Renaud and Thao-Khamsing, 2012) and could overestimate actual willingness to pay. Furthermore, experimental conditions are different from actual purchase conditions in several respects. Decisions made under such conditions do not necessarily reflect real preferences (limited rationality and behavioral bias). In this section we discuss the extent to which the results of our experiments can be applied in the case of widespread use of environmental labelling at different points of sale.

4.1. Hypothetical bias: behavior poorly understood in the literature

A hypothetical bias is likely to appear when the respondent is not in a real payment situation. The measured willingness to pay therefore overestimates individuals' real willingness to pay. The economic literature provides some indications of the extent of this phenomenon: a panel of economists from the *National Oceanic and Atmospheric Administration* (NOAA) recommends halving willingness to pay in order to reflect actual willingness to pay (NOAA, 1994 and 1996) and the meta-analysis carried out by List and Gallet (2001) even shows an average factor of three. Murphy et al. (2005) went into this work in greater detail

by testing its robustness at extreme values and only accepting studies for which hypothetical and actual willingness to pay are measured by the same disclosure method. They find an average factor of 2.6 but also show that the distribution of hypothetical factors is very right-skewed, with a median of 1.35. In other words, overestimation of average willingness to pay due to the hypothetical bias would be less than 35% for half of the studies.

In these meta-analyses, experiments carried out in the laboratory with actual payments are considered pointers to actual willingness to pay. However, measurements of willingness to pay carried out in a laboratory situation may also differ from actual willingness to pay. In particular, experiment subjects know that they can substitute the laboratory purchase with products available in their usual points of sale, which they know the price of. They can therefore reject a purchase offered as part of the experiment not because their willingness to pay is lower than the price on offer but because they plan to buy the better marketed product later. Actual prices operate a right censor on willingness to pay measured in the laboratory, which can lead to an underestimation of average willingness to pay (Harrison, 2006). The preceding results may therefore increase the extent of the hypothetical bias.

Considering the great variability of the bias between studies it is important to know more about the factors that determine the hypothetical bias, but there is only partial information on the subject. Murphy et al. (2005) show that within the stated preference methods, choice experiments tend to reduce the hypothetical bias relative to the contingent assessments in which willingness to pay is directly stated. The latter result is confirmed by a meta-analysis by Florax et al. (2005) on willingness to pay for the reduction of health and environmental risks related to the use of pesticides in agriculture. The authors find that willingness to pay revealed by hypothetical choice experiments is significantly lower than willingness to pay obtained by contingent analysis, even if it remains higher than willingness to pay obtained by disclosed preferences methods.

Furthermore, several studies tend to show that the hypothetical bias is smaller for familiar products (Johnston, 2006). Dannenberg (2009), who carries out a meta-analysis of studies measuring willingness to pay for food products without GMO, finds no significant difference between willingness to pay obtained by hypothetical choice experiments and willingness to pay obtained by experimental sales.

Finally, recent literature tends to explain the source of the hypothetical bias (see in particular Mitani and Flores, 2010, Murphy et al., 2010) but this work is still in the embryonic stage and does not allow us to anticipate to what extent the bias may vary between individuals. It is probable that this bias helps to increase the heterogeneity of estimated willingness to pay, insofar as it may affect each individual differently.

In view of these results, two characteristics of our experiments tend to limit the size of the hypothetical bias: willingness to pay is indirectly stated by choice experiments and, moreover, the choices relate to familiar products. If we rely on this literature, our study should be in line with the low range of hypothetical bias, with an order of magnitude of less than 50% for estimation of the average willingness to pay.

4.2. Comparison with willingness to pay estimated in the laboratory or under actual buying conditions

Another way of getting an idea of the size of the hypothetical bias affecting our results is to compare them to results of experiments carried out on real products with real payments. Some of them are fairly comparable to ours insofar as the measured willingness to pay concerns environment-related aspects and the study populations are from developed countries (table 5).

The first salient point in this work is the considerable variability of results, since the average willingness to pay obtained fluctuates between 10% and 250% of the product price. As this work assesses the impact of binary qualitative signals (the label is present or not), the equivalent willingness to pay in our experiments would probably correspond to two or three overall rating levels, i.e. an *average* willingness to pay in the order of 200 to 300 % in the reference version of our questionnaire. This result is a little beyond the variability range of willingness to pay obtained with real payments. Our *median* willingness to pay, in the order of 40 to 60 % for an increase of three environmental rating levels in the reference version of the questionnaire, seems consistent however with the orders of magnitude obtained with real payments (assuming for the latter that the average and the median willingness to pay are of the same order of magnitude).

Table 5 : Results of experiments with real payments for WTP relating to environmental characteristics

Authors	Publication date	Methodology	Product offered	Subject of the WTP	Average WTP		Standard deviation of WTP (euros)
					% of the product price	Euros	
Dannenberg et al.	2011	Laboratory	Chocolate bar	Non-GMO	240	0.26	0.31
Dannenberg et al.	2012	Laboratory	Oil (bottle)	Non-GMO	86	0.56	0.60
Mahé	2009	Laboratory	Chocolate bar	FT label	80	0.40	0.40
Bazoche et al.	2014	Laboratory	Golden apples (1 kg)	BA label	56	0.56	
Bazoche et al.	2014	Laboratory	Golden apples (1 kg)	Pesticide reduction	44	0.44	
Compris et al.	2011	Laboratory	Apples (1 kg)	BA label	43	0.44	
Boucherara and Combris	2009	Laboratory	Orange juice (1 L)	ecolabel	29	0.25	
Disdier and Marette	2012	Laboratory	Shrimps (100 g)	FT label	26	0.57	
Disdier and Marette	2013	Laboratory	Gherkins (1 pot)	Local product	24	0.58	
Björner et al.	2004	Experimentation	Compact detergents	Nordic Swan Ecolabel	17-29	0.44	
Disdier and Marette	2012	Laboratory	Shrimps (100 g)	BA label	21	0.50	
Björner et al.	2004	Experimentation	Toilet paper	Nordic Swan Ecolabel	13-18	0.36	
Björner et al.	2004	Experimentation	Paper towels	Nordic Swan Ecolabel	non signif.	0.14	

Note: BA: biological agriculture; FT: Fair trade

It is above all the dispersion of willingness to pay that appears important in our experiments: standard deviation there is five times higher than average willingness to pay and twenty-five times higher than median willingness to pay. In experiments carried out with real payments, the standard deviation of willingness to pay, when known, is of the same order of magnitude as the average. It therefore seems that the singularity of our results relates above all to a strong dispersion to the right of willingness to pay, connected with the choice of a log-normal distribution. This choice was, however, dictated by the data as it is this that maximizes the likelihood function.

It is possible that the hypothetical nature of the experiment allows some individuals to make their choices without taking into account product prices, hence extremes of willingness to pay would not occur in a real payment situation. Our result showing that individuals who take little care in their choices have higher willingness to pay tends to confirm this hypothesis. This type of behavior would strongly affect the dispersion and average of willingness to pay, but would have only a small effect on the median. Overall, the quality indicators of the questionnaire responses are good: average response time is only 18 minutes and a wide majority of respondents found the survey interesting and easy to answer (respectively 87% and 85% of respondents). This tends to prove that the majority of respondents have taken care over their responses.

Our experiments have another particular feature which could influence the dispersion of willingness to pay: each individual has chosen his own price range for each product family. We therefore measure willingness to pay for the product characteristics as a percentage of individual willingness to pay for the product. In the majority of studies, willingness to pay is measured as a percentage of the average price of the product, which is the same for all individuals. Our approach thus has a supplementary source of individual variability.

This particular feature of our experiment may also lead to an overestimation bias of willingness to pay. Behaviors that automatically limit choice on the basis of a price range are assumed to take place before the choice is presented to the respondents. Part of the influence of price on choices is therefore smoothed out in our measurements. More precisely, our willingness to pay refer to a situation of great diversity of references: each consumer can choose between at least three products within his own limited range of interest (regarding price but also other characteristics of the product such as flavor, design, etc.). Anyway, choice behaviors based on essential criteria contradict the very concept of willingness to pay, which

assumes that individuals compare all the product characteristics in order to make a rational choice that conforms to their preferences. This is the reason why we wanted to exclude behaviours essentially limited by price from our experiments.

4.3. Some results enable us to appreciate the extent of the measurement bias

A third way of studying the measurement bias of the willingness to pay in our experiments is to examine the internal consistency of our results, and in particular, the variability of willingness to pay according to ecolabel purchasing behavior (table 4). In particular, the median willingness to pay of consumers who never buy ecolabels (6% of the product price) constitutes an upper limit for the impact of the measurement bias on this population's median willingness to pay (unless it is assumed that over half of these individuals want the environment to be damaged). Yet the estimated median willingness to pay of the consumers who occasionally buy ecolabels is well above this limit, and they weigh half the population. This consolidates our result according to which the majority of consumers have a strictly positive willingness to pay for the environmental performance of products, and are therefore a potential target for a labelling scheme.

Willingness to pay for brand names is likely to be affected by the measurement bias of our experiments as well as environmental willingness to pay. Likewise, the averages and dispersions seem high but the medians have a reasonable order of magnitude (6 to 11% of the price of products depending on the brand). List and Gallet (2001) and Murphy et al. (2005) showed that the hypothetical bias is generally higher for public property compared with private property, even if this result is also contested (Little and Berrens, 2004). In the context of our experiment and according to this result, the bias should be smaller for willingness to pay for brand names compared with willingness to pay for the environment. However, even considering a null bias for brand names and a factor of two for the environment, median willingness to pay for products with a better environmental rating remains of the same order of magnitude as median willingness to pay for brand names. In view of the significant sums brands invest in their image and reputation, it is reasonable to think that the median willingness to pay for brand names is not zero, and therefore the median willingness to pay for the environmental rating is not either.

4.4. Purchasing behavior under real conditions does not necessarily reflect consumers' true willingness to pay

The measurement bias in willingness to pay is the primary source of the difference between our experiments and the purchasing behavior of consumers under real conditions, but there are others. Willingness to pay reflects individual preferences and it is useful to know them in order to make calculations of well-being. However, they are not perfectly expressed in a real buying context. In fact, consumers are often confronted with a very wide range of products, with numerous characteristics, such that purchase costs (reading labels and handling information) can exceed the individual's abilities or the buying issue carried out. Consumers establish routines which enable them to make a decision quickly, without perfectly optimizing each of their purchases. Automatic elimination behavior based on imperative price criteria, excluded from our experiments, falls within this category of purchase routines. At the level of extreme simplification, busy consumers choose their usual product without taking time to compare it with the others. It is only in the medium or long term that consumers will question their buying routines in order to adapt them to their preferences.

Apart from the measurement bias in willingness to pay, other biases are therefore likely to lead to behavior differences between our survey and actual conditions. Thus, a contextual bias results from the fact that the respondent's attitude at the time of responding is not the same as in a real buying situation: respondents may perhaps have more time to choose the product, they are not confronted with their usual shop and products and cannot therefore apply their usual buying routines, etc. Apart from the more considered nature of the purchases, Lusk et al. (2006) and Lusk and Norwood (2009) show that behavior in the laboratory differs from that in the field when the subject of study includes an ethical dimension. Awareness of being observed tends to lead to more virtuous behavior than under real conditions. Our experiment has limited exposure to this bias since it is an *internet* survey and not a laboratory experiment.

Only experiments conducted under real conditions can claim an absence of bias insofar as impact is concerned, but the characteristics of buyers are often poorly understood and the numbers studied are low. While the internal validity of these experiments is good, their external validity may be bad. In particular, small scale experiments of short duration generally do not enable us to assess the impact of a long term scheme at the European scale.

4.5. Conclusions about the scope of our results under real conditions

Due to the virtual nature of the choice experiments carried out in our survey, our results include a hypothetical bias which probably leads to an overestimation of the average and of the dispersion of willingness to pay. This bias is perhaps increased by the elimination, in our experiments, of behavior consisting of pre-selection of products based only on a price range. Concerning the factors available in the literature, the hypothetical bias should be quite low with respect to our estimations of median willingness to pay. Nevertheless, the measurement bias is unlikely to challenge the result according to which a majority of consumers have a positive WTP for the environmental performance of products.

Apart from the hypothetical bias, our results may differ from behavior that would prevail in the real world situation. The willingness to pay that we measure illustrates the potential of labelling when it is implemented in a standard, highly visible way over a relatively long period of time. Environmental labels used in the survey are homogeneous and placed in a prominent position on the products. This "ideal" scope is even more marked as respondents are given prior information about the nature of the label and what it represents (appendix 3). Our survey therefore fits into a context in which labelling would be the subject of publicity campaigns and in which consumers would have had time to familiarize themselves with it. Furthermore, the survey tends to favor consideration of the environmental label by a small number of characteristics differentiating the products. In a real-world situation, packaging, smell or visual differences may deflect the consumer's attention away from environmental footprints.

Finally, the survey resembles an on-line purchasing platform which reproduces a considered type of shopping. Yet, in real-world situations, trade-offs are dependent on habit and the limited amount of time allocated to shopping, which lead to routine behavior such as repeatedly buying the same product. We may nevertheless assume that this behavior is open to questioning by consumers in the long term. Experience gained, particularly from energy labels on household appliances, actually proves that acceptance of a new labelling scheme by the general public is a process requiring several years.

Despite all these limitations, the ecolabel purchasing behavior stated by consumers is very closely related to the willingness to pay revealed in the choice experiments. This result tends to prove that our virtual willingness to pay contains relevant information for predicting purchasing behavior under real conditions. Moreover, the hypothetical bias or behavioral biases do not need to vary greatly between the different versions of the questionnaire, such that comparisons between versions remain legitimate. In particular, the superiority of mandatory labelling over voluntary labelling is not called into doubt.

Conclusion

The development of a survey with hypothetical choice experiments provides a wealth of information about consumer behavior in relation to a scheme of environmental labelling displayed on consumer products. In fact, consumers' willingness to pay appears significant with respect to products with a higher environmental rating. It particularly seems to exceed willingness to pay associated with brands with high commercial potential. The buying habits of the majority of consumers appear to be open to influence by an environmental labelling scheme, well beyond the niche markets often associated with environmental logos. Furthermore, a mandatory environmental label would be more effective than a voluntary label for guiding buying behavior: consumers pay more heed to the environmental characteristics when they are displayed on all products. In addition, unlabelled products are perceived relatively well by consumers, so well in fact that a voluntary labelling scheme would not tend to be spread spontaneously.

These results are subject to bias mainly due to the virtual nature of purchases made during the survey. However, some results tend to show that this bias is not large enough to qualitatively challenge our results.

The work presented here is far from exploiting the full extent of the survey. In particular, further work could explore the factors determining consumers' willingness to pay, through a typology of their attitudes towards the environment. Researchers wishing to access the data for their own work can obtain them for free[17].

[17] Researchers may request data by email from: Ernr.Seei.Cgdd@developpement-durable.gouv.fr.

References

Ardilly, P. (2006), « Les techniques de sondage », TECHNIP, chapter II-6.

Bigot, R., P. Croutte and F. Recours (2010), « Enquêtes en ligne : peut-on extrapoler les comportements et les opinions des internautes à la population générale ? », CREDOC, Cahier de Recherche n°273, December 2010.

Bazoche, P., P. Combris, E. Giraud-Héraud, A. S. Pinto, F. Bunte and E. Tsakiridou (2014), "Willingness to pay for pesticide reduction in the EU: nothing but organic?", European Review of Agricultural Economics, vol 41, pp 87-109.

Bernard, Y., L. Bertrandias and L. El Gaaïed (2012), « L'efficacité de l'étiquetage environnemental : résultats d'une étude quasiexpérimentale exploratoire », forthcoming working document (2012), unpublished reference.

Bjorner, T. B., L. G. Hansen and C. S. Russell (2004), "Environmental labeling and consumers' choice – an empirical analysis of the effect of the Nordic Swan", Journal of Environmental Economics and Management 47:411-434.

Bleda, M., and M. Valente (2009), "Graded eco-labels: A demand-oriented approach to reduce pollution", Technological Forecasting & Social Change 76, 512–524, 2009.

Bougherara, D., and P. Combris (2009), "Eco-labelled food products: what are consumers paying for?", European Review of Agricultural Economics, 36 (3), 321-341.

Bougherara, D., and Piguet (2008), « Marchés avec coûts d'information sur la qualité des biens : une application aux produits écolabellisés », Economie et Prévision n°182.

Brecard, D. (2014), "Consumer Confusion over the Profusion of Eco-labels: Lessons from a Double differentiation Model", Resource Energy Economics 37:64-84.

Cambell, D. (2014) "Elimination by aspects in discrete choice experiments: implications of not accounting for dominant attributes"

Cantillo, V. and J. Dios Ortuzar (2004) "A semi-compensatory discrete choice model with explicit attribute thresholds of perception"

Couper, M. P. (2000), "Web Surveys – A Review of Issues and Approaches", Public Opinion Quarterly, vol. 64, pp 464–494.

Combris P., Seabra Pinto A., Bazoche, P., Berjano, M., Giraud-Heraud, E., Hanus C. and Maia R. (2011), "Consumer's willingness to pay for reduced pesticides use on production of fresh and processed apples". Working document.

Dannenberg, A. (2009), "The dispersion and development of consumer preferences for genetically modified food – A meta-analysis", Ecological Economics 68: 2182-2192

Dannenberg, A., S. Scatasta and B. Sturm (2011), "Mandatory versus voluntary labelling of genetically modified food: evidence from an economic experiment", Agricultural Economics 42:373–386.

De Palma, T. (1987), « Les Modèles de choix discrets ». Annales d'économie et de statistique N°9.

Disdier, A-C., and S. Marette (2012), "Taxes, minimum-quality standards and/or product labeling to improve environmental quality and welfare: Experiments can provide answers", Journal of Regulatory Economics, vol.41, n° 3. pp. 337-357.

Disdier, A-C., and S. Marette (2013), "Globalisation Issues and Consumers' Purchase Decisions for Food Products: Evidence from a Lab Experiment", European Review of Agricultural Economics, vol. 40, n° 1. pp. 23–44.

Florax, R. J. G. M., C. M. Travisi and P. Nijkamp (2005), "A meta-analysis of the willingness to pay for reductions in pesticide risk exposure", European Review of Agricultural Economics Vol 32:441-467.

Frippiat, D. and N. Marquis (2010), « Les enquêtes par Internet en sciences sociales : un état des lieux », Population – vol.65, pp 309-338.

Harrison, G. W. (2006), "Experimental Evidence on Alternative Environmental Valuation Methods ", Environmental and Resource Economics, Vol. 34, Issue 1, pp 125-162.

Johnston R. J. (2006), "Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum", Journal of Environmental Economics and Management 52:469-481.

List, J. A. and C. A. Gallet (2001), "What experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis", Environmental and Resource Economics 20:241-254.

Little, J. and R. Berrens (2004), "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta–Analysis, Economics Bulletin Vol. 3, No. 6., pp. 1-13

Lusk, J. L., J. R. Pruitt and B. Norwood (2006), "External Validity of a Framed Field Experiment", Economics Letters 93:285-290.

Lusk and Norwood (2009), "Bridging the Gap between Laboratory Experiments and Naturally Occurring Markets: An Inferred Valuation Method", Journal of Environmental Economics and Management 58:236-250.

Mahé, T. (2009), « Préférences des consommateurs pour des aliments avec des labels sociaux et environnementaux. Approches empiriques, expérimentales et comportementales », doctoral thesis in economic sciences at Pierre Mendès University, France.

McFadden, D. and K. Train (2000) "Mixed MNL for discrete response". Journal of Applied Econometrics 15, 447-470.

Mitani and Flores (2014), "Hypothetical Bias Reconsidered: Payment and Provision Uncertainties in a Threshold Provision Mechanism", Environmental and Resource Economics Volume 59, Issue 3, pp 433-454.

Murphy J. J., P. G. Allen., T. H. Stevens and D. Weatherhead (2005), "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation", Environment and Resource Economics 30:313-325, 2005.

National Oceanic and Atmospheric Administration (1994), "Natural resource damage assessment: proposed rules", Federal Register 59:23098-23111.

OECD (2013), « Caractérisation des dispositifs d'éco-étiquetage et d'information environnementale », OECD, Paris, France.

Ruffieux, B. (coord.), C. Gomy, A. Lacroix and L. Muller (2014) - « Projet Lab2Green. Une étude expérimentale de l'impact sur les choix des consommateurs de différents systèmes d'évaluation environnemental apposés en face avant des aliments. Rapport scientifique » - Paris : Ministry of Ecology, Sustainable Development and Energy, p.62 (MOVIDA Programme)

Scarpa, R., M. Thiene, and K. Train (2008), "Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps". American Journal of Applied Economics 90, 94-1010.

Spence, A. M. (1975), "Monopoly, quality and regulation", Bell Journal of Economics 6:217-429.

Train, Kenneth E.(2003), "Discrete choice methods with simulation", Cambridge University Press, second edition, 2003.

Train, K., M. Weeks (2005). "Discrete choice models in preference space and willingness-to-pay space". Application of simulation methods in environmental and resource economics.

Tversky, A. (1972) "Elimination by aspects: a theory of choice". Psychological Review 79, 281-299

Vanclay J. K., J. Shortiss, S. Aulsebrook., A. M. Gillespie, B. C. Howell, R. Johanni, M. J. Maher, K. M. Mitchell, M. D. Stewart and J. Yates (2011), "Customer Response to Carbon Labelling of Groceries", J Consum Policy 34:153-160, 2011.

Weiss, K. and F. Girandola (edited by) (2010). « Psychologie et développement durable », Paris, In Press.

Zago, A. M. and D. Pick (2004), "Labelling policies in food markets: Private incentives, public intervention and welfare effects", Journal of agricultural and resource economics 29:150-176.

Appendices

Appendix 1: Linearity of preference for environmental rating

One of the aims of the study is to measure the additional cost the consumer is prepared to assume in order to buy a product of better environmental quality, more specifically for a differential of an overall rating ("A" rather than "B" rather than "C", etc.). The estimated model assumes that the effect on utility of a higher environmental rating is linear.

We test this hypothesis by allowing distinct willingness to pay for each value of the environmental rating, the lowest value being taken as the reference (value "E"). This enhancement of the model was carried out on a simplified version with respect to willingness to pay for well-known brands, which are assumed to be the same in the three product families. The expression of consumer utility is modified as follows:

$$U_{itj} = \alpha_{i1} 1_{j=1} + \alpha_{i2} 1_{j=2} + \lambda_i \left[WTP_{ratingA,i} 1_{Rating=A,itj} + WTP_{ratingB,i} 1_{Rating=B,itj} + WTP_{ratingC,i} 1_{NRating=C,itj} + WTP_{ratingD,i} 1_{Rating=D,itj} + WTP_{Brand,i} 1_{Brand,itj} + WTP_{volume,i} \log(Volume_{itj}) - \log(p_{itj}) \right] + \varepsilon_{itj}$$

The estimation results show a remarkable linearity of median willingness to pay for the various environmental rating values (graph A4). In addition, the normal distributions underlying the WTP are correlated at more than 98% (table A4), which means that individuals with a higher WTP for the "A" rating also have a higher WTP for "B", "C" and "D" ratings. Linearity is therefore relevant at the individual level and not just at the median level.

Graph A4: Median willingness to pay for the different environmental rating values

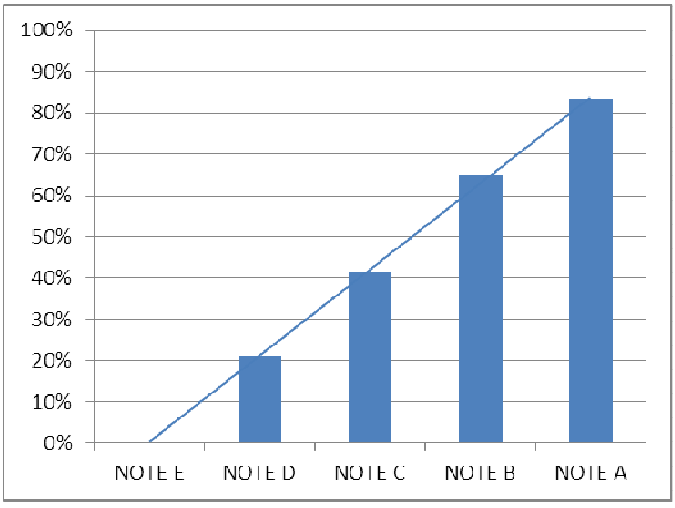


Table A4 : Extract of the estimated correlations matrix for normal distributions underlying the individual parameters

	$CAP_{NOTE\ A}$	$CAP_{NOTE\ B}$	$CAP_{NOTE\ C}$	$CAP_{NOTE\ D}$
$CAP_{NOTE\ A}$	100 %	99.7% (0.1%)	99.0% (0.2%)	98.1% (0.3%)
$CAP_{NOTE\ B}$		100%	99.4% (0.2%)	98.1% (0.3%)
$CAP_{NOTE\ C}$			100%	98.0% (0.6%)
$CAP_{NOTE\ D}$				100%

Appendix 2: Correlation matrices

Table A2. Estimated correlation matrix for normal distributions underlying the individual parameters vector in the reference version of the questionnaire (standard deviation in brackets)

	α_1	α_2	λ	WTP_{rating}	$WTP_{Washing\_powder\_brand}$	WTP_{Jeans_Brand}	$WTP_{Yoghurt\_Brand}$	WTP_{volume}
α_1	100 %	32% (16%)	16% (10%)	-22% (9%)	-21% (13%)	-9% (11%)	2% (12%)	-4% (18%)
α_2		1	-12% (17%)	-32% (19%)	-14% (14%)	26% (16%)	30% (16%)	1% (16%)
λ			1	-77% (3%)	-41% (6%)	-51% (5%)	-42% (7%)	-56% (8%)
WTP_{rating}				1	24% (6%)	25% (6%)	20% (7%)	33% (8%)
$WTP_{Washing\_powder\_brand}$					1	49% (8%)	64% (6%)	36% (10%)
WTP_{Jeans_Brand}						1	62% (7%)	24% (9%)
$WTP_{Yoghurt\_Brand}$							1	19% (9%)
WTP_{volume}								1

Appendix 3: Presenting the environmental label to the respondents

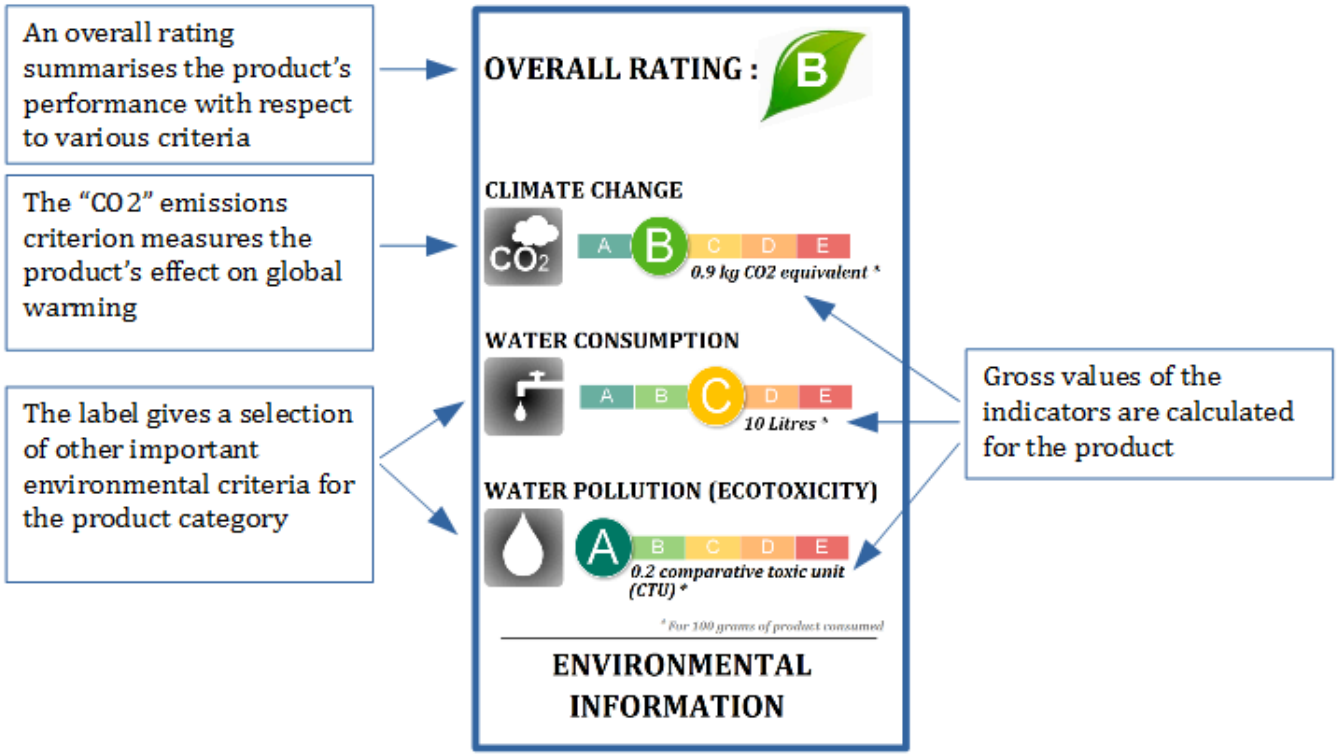
Before taking the rest of the questionnaire, here is a presentation of a new scheme planned by the Ministry of Ecology, Sustainable Development and Energy: ENVIRONMENTAL LABELLING of mass consumption products.

This labelling scheme will present the environmental performance of products and will thus enable consumers to compare them.

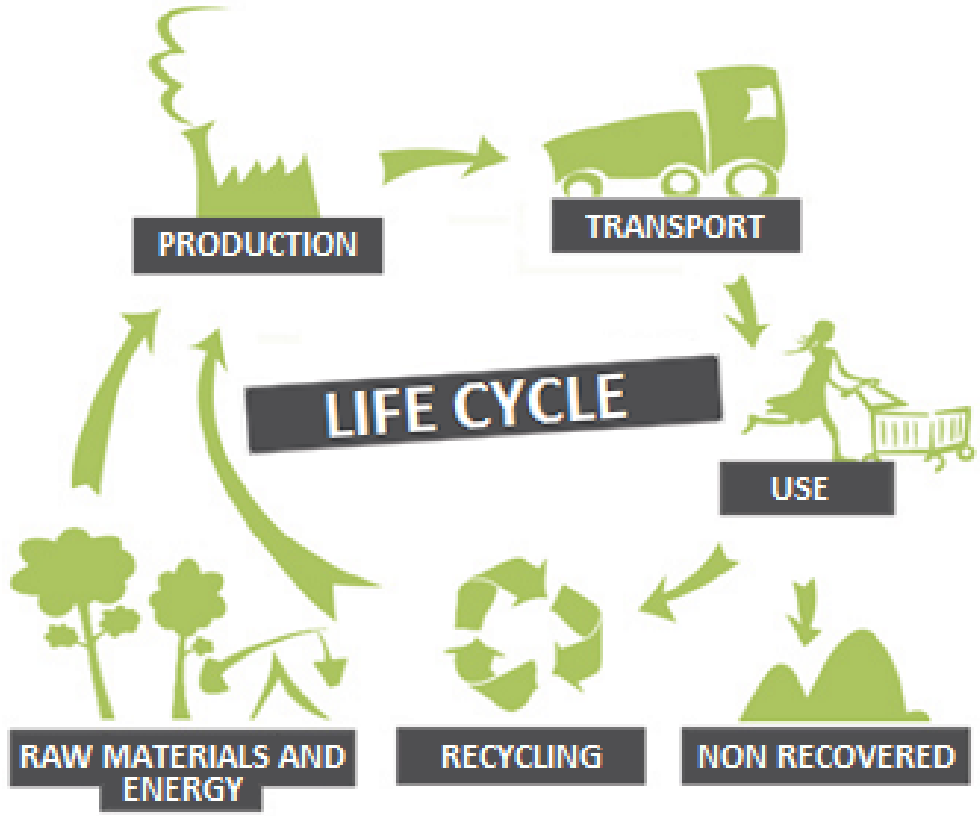
Environmental performance is quantified information about the main environmental impacts of the products, calculated according to an official method, with guaranteed independent oversight.

Here is an example:

Example of an environmental label



Indicators take into account the whole life cycle of the products: extraction of raw materials, production, distribution, use and end of life (including recycling).



Appendix 4: Education pages of the questionnaire for yoghurt product family

Here is some information about the environmental impact of yoghurt

Climate change

Greenhouse gas emissions cause climate disruption for the whole planet. Milk production for the manufacture of yoghurt contributes to these emissions.

- Cows produce a greenhouse gas (methane) during digestion (belching and flatulence).
- Spreading manure to fertilise grassland gives off a greenhouse gas (nitrous oxide).

Water pollution (eutrophication)

Some products promote the proliferation of algae in aquatic environments and put ecosystems at risk.

- Fertilisation of grassland for growing hay and providing pasture for cows uses manure which is partly removed by rain and carried away in rivers.

Dairy cow

Photograph by USDA – ARS

Loss of biodiversity

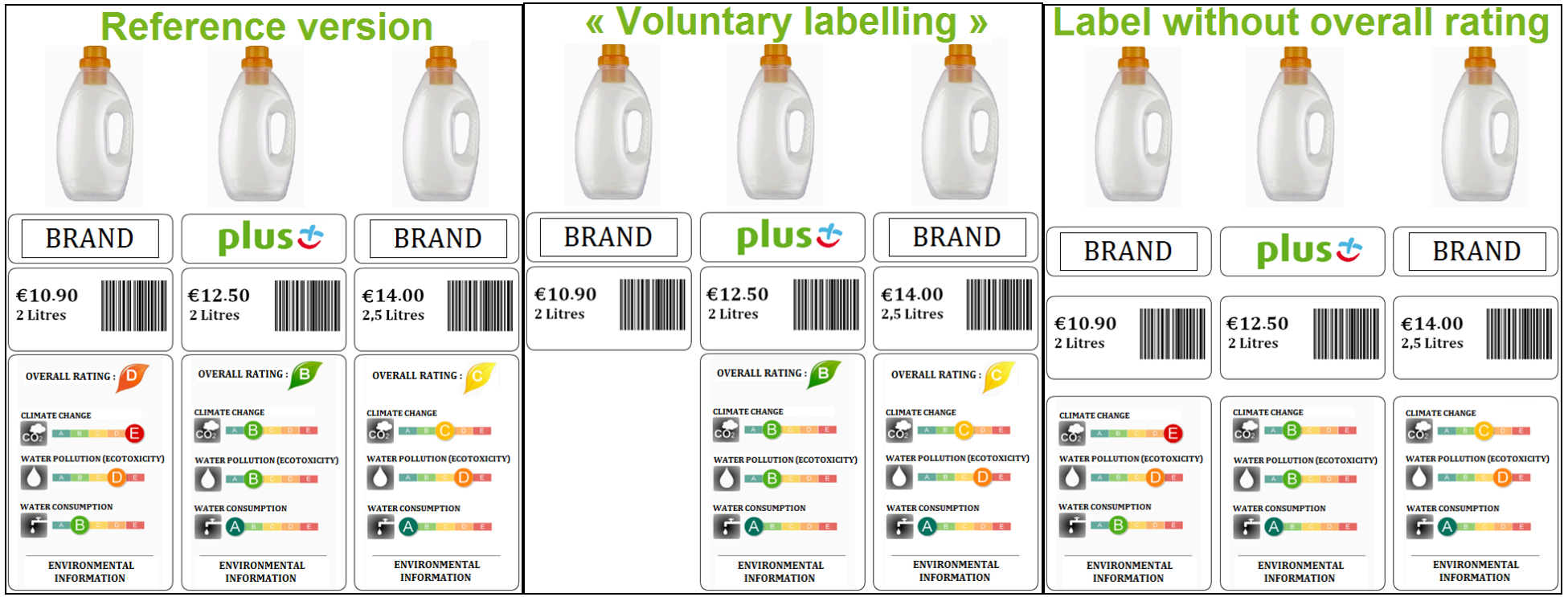
Some species and some ecosystems are becoming rare and are threatened with extinction.

The increasing intensification of dairy farming is speeding up this loss of biodiversity.

- The trampling and grazing of grassland by cows prevents some plant species from developing properly.
- The early cutting of grass for hay production prevents some animal and plant species from fully completing their biological cycle of reproduction

To reduce the environmental impact of your yoghurt consumption, you can choose the products with the best environmental information.

Appendix 5: Examples of choice experiments



Commissariat général au développement durable
Service de l'économie, de l'évaluation et de l'intégration du développement durable
Tour Séquoia
92055 La Défense cedex
France
Tél : 33(1).40.81.21.22

You may connect to the following website to view this publication:
http://www.developpement-durable.gouv.fr/

The impact of environmental labelling on consumer choices: lessons from a large-sample choice experiment

The aim of environmental labelling is to educate consumers about the environmental impact of the products they buy. Unlike ecolabels, which are manifested by the presence of a logo on consumer products, the principle of environmental labelling is to display quantitative information on the environmental footprint of the products, in the form of a graduated scale for example. This study analyses the impact of such a scheme on consumer choices, using an original survey of 5,000 respondents with hypothetical choice experiments. We estimate consumers' willingness to pay for better environmental quality from their product choices. Our results show a high level of sensitivity to environmental quality with considerable heterogeneity of preferences between consumers. The median of willingness to pay for more environmentally friendly products would be nearly double the willingness to pay for well-known brands. Our results also show that environmental labelling is likely to reach a larger target population than ecolabels. Finally, the effectiveness of the scheme varies depending on whether it is mandatory or voluntary: when the label is not mandatory the consumer tends to consider unlabelled products as being of average environmental quality, leading producers to only reveal their environmental performance when it is above average. Environmental labelling then comes closer to ecolabels.

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère
de l'Environnement,
de l'Énergie
et de la Mer

Legal Deposit : May 2016
ISSN : 2102 - 4723

Service de l'économie, de l'évaluation et de l'intégration du développement durable