Testing CO2/Car labelling options and consumer information

FINAL REPORT

Cristiano Codagnone[$], Francesco Bogliacino[*], and Giuseppe Veltri[#]

[$]Università degli Studi di Milano, and Universitat Oberta de Catalunya (UOC)

[*]Fundación Universitaria "Konrad Lorenz"

[#]University of East Anglia

21 June 2013

Client:
European Commission
DG Climate Action

Client Contact:
Cristiano Codagnone (*)
MediaTic, Roc Boronat 117
08018 Barcelona – Spain
Tel: +34934505351
Mob: +34676837591

Contract Reference:
071201/2012/625659/CLIMA.C.2 implementing Framework Contract No EAHC / 2011 / CP / 01/LSE

Consortium LSE & Partner:
London School of Economics
Tech4i2 Limited
Block de Ideas
Universitat Oberta de Catalunya (UOC)

Confidentiality, copyright & reproduction:
This report is the Copyright of European Commission and has been prepared by the LSE & Partner Consortium. The contents of this report may not be reproduced in whole or in part, nor passed to any organisation or person without the specific prior written permission of European Commission.

Authors:
C. Codagnone, F. Bogliacino, Giuseppe Veltri

Approved by:
G. Gaskell, LSE

Date:
June 21 2013

Signed:

This report was produced in the frame of a contract with the Executive Agency for Health and Consumers (EAHC) acting under its mandate from the European Commission. The content of this report represents the views of the LSE & Partners Consortium and is its sole responsibility; it can in no way be taken to reflect the views of the European Commission and/or EAHC or any other body of the European Union. The European Commission and/or EAHC do not guarantee the accuracy of the data included in this report, nor do they accept responsibility for any use made by third parties thereof.

Executive Summary

This is the final report of the study "Testing CO2/Car Labelling Options And Consumer Information". The overall aim of the study was to test in experimental settings the effectiveness of possible new variants of car eco-labels and promotional material. In order to accomplish this overall goal the study included the following tasks:

- **Task 1: Preparatory Phase**. This comprised: a) a review of the literature; and b) the conducting of a preliminary survey in 10 countries (BE=Belgium;DE= Germany; FR= France; IT= Italy; NL= Netherlands; PL= Poland; RO= Romania; ES= Spain; SE= Sweden; UK= United Kingdom) with a total sample of N= 8000 respondents (800 per country);
- **Task 2: Laboratory experiment testing labels**. A laboratory experiment was conducted at the LSE Behavioural Lab (N=405 respondents);
- **Task 3**: **Online experiment re-testing labels and testing promotional material.** This was realised conducting an online experiment in 10 countries (BE=Belgium;DE= Germany; FR= France; IT= Italy; NL= Netherlands; PL= Poland; RO= Romania; ES= Spain; SE= Sweden; UK= United Kingdom) with a total sample of N= 8000 respondents (800 per country);
- **Task 4**: **Analysis and Recommendations**. We performed econometric and multivariate statistical analysis on the data from the three empirical components listed above and we triangulated the findings to extract policy implications and recommendations.

Objectives

The client, DG CLIMA, aimed to test in experimental settings:

- **The effectiveness of car labels**. Test the effectiveness of alternative car labels, with specific reference to their content and layout, in terms of informing potential car buyers about - and raising awareness of - CO2 emissions, fuel efficiency, and running costs of new cars;
- **The effectiveness of mandatory information in promotional material**. Test the effectiveness of mandatory information in promotional material in terms of informing potential car buyers about, and raising awareness of, CO2 emissions, and fuel efficiency of vehicles.

The preliminary survey, as instrumental and preparatory to the two experimental tests, aimed at: a) exploring the car-purchase process and the specific (main attributes such as price, safety, fuel efficiency, etc.) and contextual (self-reported attitudes and intentions) parameters shaping it; and b) investigating the awareness among consumers about the environmental impact of car usage.

At the end of the survey, however, the sample was randomly split into four sub-samples (split ballot technique) each allocated to see the same car (a VW Polo) but matched to four different simplified labels. The labels contained only the CO2 emissions using four different classification systems:

1. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the Relative Classification system (i.e. a vehicle is rated compared to vehicles belonging to the same class);
2. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the Absolute Classification system (i.e. a vehicle is rated compared to vehicles from all classes);
3. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the Combined Classification system (i.e. a combination of the absolute and relative rating);
4. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the German classification system (a sub-variant of the relative classification).

Then they were asked four questions about the car model, they had just seen associated with one of the four classification systems, to test the extent to which such systems had a noticeability effect:

- Q37a: 'How environmental friendly do you think this car is?'
- Q37b: 'How fuel efficient do you think this car is?'
- Q37c: 'How environmentally-friendly do you think this car is compared to other similar type of cars?'
- Q37d: 'How environmentally friendly do you think this car is compared to the car/s in your household?'

The split ballot was only an exploratory test, sufficient to shape the design choices of the laboratory experiment but not to provide a conclusive judgement on the relative effectiveness of the four classification systems.

In the laboratory and in the online experiments we tested the effectiveness of the following pieces of information and/or their layout:

For Labels

- **Standard elements**. We mean elements that, though in different formats, are already contained in most of the labels and promotional materials currently in use, such as the different CO2 classification systems (absolute: car rated compared to cars from all

classes; relative: car rated compared to cars belonging to the same class; combined: a combination of absolute and relative rating) and their format (text only, in graphic format using either a vertical or a horizontal layout);

- **Additional elements**. We mean elements that either are not included in currently used labels and promotional material or are included only in a few countries. Elements such as :
  - Information about running costs;[1]
  - Information about taxation related to CO2 emissions;
  - Information about the savings lost on fuel that buying a car X implies, if compared to the most efficient car in the class;

For Promotional material

- **General format:** concerning the layout by which information about CO2 emissions is provided;
- **Additional Elements:** concerning mostly the layout by which information about CO2 emissions is combined with information on running costs;
- **Web link:** the presence or absence of this element.

These experimental conditions were administered to participants by way of visual stimuli containing different combinations of the pieces of information (and/or their layout) listed above.

Therefore, we report the results of the effect of these items of information as separate individual treatments in a main effect design. We did not test labels or promotional material as holistic treatments, as this would have required a full factorial design testing of all the possible interaction effects among pieces of information and their formats. The readers can find a smaller selection of the visual stimuli used in the Appendix of this report (chapter six), a larger one in the Technical Compendium (Annex II delivered separately), and a full set in Annex IV (also delivered separately).

Rationale, design, and analysis

Eco-behaviour concerns attributes that cannot be as easily verified by consumers as elements directly visible or experienced; such as for instance colour, road handling, acceleration, etc. The colour of a car or its acceleration power, not only can be directly experienced, but also directly tap into consumers' emotions and intuitive thinking, a fact

[1] We use this general expression to refer to the running costs in terms of fuel only and we tested it as such without adding the other elements of maintenance costs (i.e. tires, oil, other consumables, insurance, etc.). We have agreed this choice with DG CLIMA main on the basis of two reasons, one substantive and one pragmatic. First, the piloting of the experiment showed that most people intuitively associated running costs to fuel. Second, for the online experiment in 10 countries it was beyond feasibility to find comparable and consensual measures of the other maintenance cost elements.

that manufacturer's brand campaigns saliently leverage. Eco-friendliness and fuel efficiency are indeed related and could be objectively verified, but information about these two aspects is either provided in an unclear and unfriendly way or is simply not processed by consumers due to what behavioural scientists call heuristics and biases. The review of the literature presented in the Technical Compendium (chapter 2) fully confirms this and other related aspects such as: the presence of an 'attitudes-action gap' (consumers are aware of the environmental impact of cars but this does not translate into eco-friendly car purchases), inconsistent preferences (self-reported importance of fuel economy and then purchases of fuel inefficient models), a widespread level of eco-ignorance and lack of understanding of the connection between CO2 emissions and fuel efficiency.

The perspectives of behavioural economics can explain the inconsistencies in preference, eco-ignorance, and attitude-action gap and help policy makers devise measures contributing to induce better choices both for the environment (CO2 emissions) and for the budget of consumers (fuel efficiency). According to the well known synthesis, made by Nobel Prize winners Daniel Kahneman (2011), consumers cognitive bias depends on the interaction between the automatic and affective 'hot cognition' of what he calls System 1 and the more reflective and cool cognitive mood of System 2. The manifesto of the behavioural approach to consumer policy, Thaler and Sunstein (2008), illustrates the mechanisms (called nudges) by which policy makers can de-bias consumers' mental shortcuts and inertia.

We have applied this perspective and tested 'nudges' that, by better framing of existing information elements and/or by including clear and easy to understand new elements, aimed to activate reflective and cool cognition in consumers and induce more eco-friendly and fuel-efficient choices. In order to measure the effectiveness of the proposed nudges we used both behavioural variables, registered automatically through the choices the participants made in a simulated purchase experimental task and self reported cognitive responses to various questions measuring cognitive effects (noticeability; and comprehension and recall)

In the online experiment, before participants were randomly allocated to the experimental task, we also asked questions about existing labels that we use to model the relationships among a set of variables that explain the usage of labels.

The two experiments were designed as a randomised control trial, with a classic between subject design, whereby participants are allocated to a treatment (nudge) or to a control condition (placebo). In both we performed multiple randomisation protocols so that participants were randomly allocated to conditions. Whenever participants were exposed

to more than one experimental condition the order by which they were exposed to them was also randomised, as were the visual stimuli containing the treatments, and the cars to be selected for the simulated purchases. The cars were randomly selected from a database of 470 cars (including all relevant information and the image of the cars) that we constructed ad hoc for this study. This database covers exhaustively the different class sizes and the three different types of engine (conventional, electric, and hybrid).

For the preliminary survey and the online experiment (conducted in the same 10 countries but in different periods and with different sample) we have a representative random sample of 8000 participants (800 per country). For the laboratory experiment we used a convenience sample of 400 participants recruited from the panel managed by LSE Behavioural Laboratory.

Given the design used, we recover through regression analysis the difference in means between treated groups and control groups with respect to the response variable used (behavioural measure of willingness to pay, self-reported and constructed cognitive measures), which correspond to causal effect of the treatment (nudge) considered. We first run Ordinary Least Square (OLS) and then perform all required technical checks (using alternative specifications of regression analysis). We always control also for age, gender, and country. The results we present in this report are fully robust to all the controls.

Answers to the pre-treatment questionnaire have been processed through a Structural Equation Model in order to explain what shapes the usage of currently existing labels.

In the case of the preliminary survey we have produced standard descriptive statistics (reported in full in Annex III delivered separately) and performed an analysis of variance, with respect to the split ballot, exploring the cognitive effects (in this case only in terms of noticeability) of the different CO2 classification systems.

Main findings of the preliminary survey

The following are the most noteworthy findings of our preliminary survey;

- **Awareness:** Our survey confirms that Europeans are aware and not indifferent to the environmental problems caused by the pollution produced by vehicles:
  - Whereas currently only about 1.9% of the sample possess either a hybrid (1.8%) or an electric (0.1%) vehicle, as many as almost 33% of respondents say they will buy electric or hybrid as their next car;
  - More than 50% of the respondents think that more than 40% of the greenhouse effect is attributable to car pollution;

- A clear majority of respondents is aware that their behaviour has an impact on the environment and thinks that they can make a change with their actions;

- **Attitudes-action gap:** our results also confirm the gap observed in literature between self-reported attitudes/intentions and actual behaviours. This gap can be appreciated comparing the findings about awareness, reported under the previous point, with the following ones:
  - Environmental concerns come after 10 other main attributes (price, safety, performance, etc.) in terms of importance in influencing car purchase decisions;
  - When asked a question about preferences clearly related to their actual behaviour, the majority of respondents revealed a preference for their own private car as compared to alternative means of transportation for daily commuting;
  - Consumers first select a class of vehicles, and only when they narrow down to choosing a model they may then take into considerations eco-friendly parameters;

- **Moderate familiarity with existing labels:**
  - More than half of respondents report not being very familiar with labels;
  - 40% disagree with the statement that they are easily recognisable;
  - 44.5% agree that car labels are unfamiliar to them;
  - Many also misunderstand environmental labels as symbolising product reliability.

The preliminary exploration on the relative effectiveness of the CO2 classification systems, through the split ballot, suggests that the relative classification system is more effective in eliciting noticeability effects than the German and the combined systems.

Main findings of the two experiments and interpretation

Both the laboratory experiment and the online experiment tested variants of information elements (and/or their layout) concerning labels, whereas the latter also tested items concerning promotional material.

Although conducted under different conditions and with different samples, for the labels variant the two experiments fed one into the other and it makes sense to provide some general common findings and interpretation. More granular and distinct analysis of the results is presented in narrative form in this report (in chapter 3 and chapter 4) and with more technical details in the Technical Compendium (chapter 4 and 5).

Labels:

- **Non systematic findings**: this means we do not get an unequivocal picture of what works and what doesn't. Although we find some statistically significant effects both in

the laboratory experiment and in the online experiment, they are not consistent across treatments, measures, and engine types, but somehow scattered across them
    - The effects are more common in terms of noticeability;
    - Effects on willingness to pay are evident only for electric and hybrid cars;
    - In brief we do not get a clear direction of effectiveness that allows discriminating between what works and what does not work;

But we have also some more conclusive findings:

- **CO2 classification system**: if we consider the cognitive processing of emission information, the absolute system overall works better than others and is a recommended and conclusive choice;
- **Nudges related to fuel economy work better than emissions related nudges**
    - Several items across the two experiments perform fairly well, for instance:
        - Information on lost savings on fuel;
        - Running costs in different formats;
        - Information on electricity consumption (electric cars only);
        - Information about consuption in the two formats specific to hybrid car;
    - If we triangulate the findings of the laboratory experiment, those of the online experiment for the labels,and also the results for the promotional material, we can conclude that running costs (in both the per mile/km format and in the per 5 years format) is the relatively more effective nudge among all those we have tested;

Promotional materials:

- **Non systematic but larger set of significant findings**: although findings are still not fully consistent and systematic, for promotional material we find a larger number of statistically significant effects as compared to the nudges concerning the labels;
- **Significant effects across variables**: we find more statistically significant effects on the comprehensive measurement scales used;
- **Most effective treatments**: the two more effective nudges are:
    - The use of a graphic illustration of CO2 emissions;
    - The use of a larger element indicating running costs per 5 years;
- The **web link** has contrasting effects and our results are not conclusive as to the overall effectiveness of this nudge.

Whereas we condensed above the findings from the experiment in plain and neutral fashion, below we provide our tentative interpretation of such findings adding our own hypotheses.

Our main hypothesis is that the nudges we tested were to some extent overloaded with information and had mixed effects.

In the case of eco-labels there is no possibility to use the very effective nudge represented by changing default options in consumer purchase settings. For reasons extensively explained in the report using nudges tapping directly into consumers' emotions was also not feasible. Therefore, the variants of information elements, and/or their layout, that we tested fall into the category of nudges that, by better framing existing information elements and/or by including clear and easy to understand new elements, aim to activate reflective and cool cognition in consumers. Our interpretation of the fact that we find some effects, but not in systematic fashion, is because the information elements we tested as nudges do not affect directly System 1 (hot cognition) and do not fully activate cool cognition of System 2. This latter aspect is probably due to the fact that there was too much information to process cognitively. Two indirect forms of interpretative reasoning can confirm this.

First, cognitive measures with respect to the fuel economy show somewhat better results for the impact of electricity consumption, running costs per mile (conventional and hybrid engines) and running cost per 5 years in the case of the electric cars. In this case we can propose the following interpretation in line with the main hypothesis presented here. The information is much starker and/or might benefit from the subconscious assumptions about the electric car; as a result confirmatory reasoning becomes less cognitively demanding.

Second, promotional material visual stimuli are simpler than their label counterparts and, thus, may be easier to process and use even without a full activation of System 2. This may explain why we find a larger number of statistically significant results for promotional material treatments compared to labels' treatments.

Additionally, the higher effectiveness of promotional material could be explained by two other factors. First, within promotional material there is more variation both among the treatments, and between them and the control condition. The labels are constrained by a common structure that makes treatment more similar to each other and, especially, not starkly different from the control condition. Second, it is quite possible that promotional material may represent for consumers a more familiar form in a purchase context than the labels. This is an important element that can be related to the issue of familiarity, trust, comprehension, and usage of labels to which we turn below.

Explaining labels usage

The model we run, to explain the usage of current labels, processed the answers provided by all 8000 participants of the online experiment before they performed the experimental tasks and were exposed to the treatments (nudges). This means, on the one hand, that the results of this model cannot be directly related to the results of the regression analysis discussed above. On the other hand, it also means that the answers were not affected by exposure to such nudges and, thus, reflect attitudes and experiences formed as a result of real life exposure (or lack of) to currently used labels. As such, the results of this explanatory model can be indirectly triangulated with those of the regression analysis and help us contextualise the fact that we do not find systematic and consistent results.

The main result of our explanatory models is that labels usage is mostly explained by the familiarity respondents have with existing labels, and on how much they trust them. The comprehension of the label has a minimal direct effect on usage, and its effect is mostly mediated through familiarity and trust.

It should not have escaped the attentive reader that the comprehension of labels is the most important cognitive dimension, which could be impacted through policy nudges such as those we tested both in the laboratory experiment and in the online experiment. Our model tells us that comprehension has almost no direct impact on label usage, but all its impact is mediated through familiarity and trust. Therefore, providing the best and most effective nudges in the short run will not impact label usage (which in turn is expected to increase eco-friendly purchase and behaviour in using means of transportation), but will gradually impact positively on familiarity and trust and through them the usage of labels. This means that policy interventions should be provided steadily and can be expected to have the desired effect over time. As a corollary, it is also possible to see the non-systematic nature of the effects found as a result of lack of familiarity, in light of the more robust results in the domain of promotional material (that are more familiar to consumers).

Conclusions and recommendations

Increase familiarity and trust with easy to understand labels

The explanatory model of labels usage clearly tells us that, at the general population level, it is not realistic to influence in the short run the use of eco-labels (and subsequently, the behaviour) by only providing better nudges that are easier to process and understand. The direct effect of label comprehension on label usage is very limited, whereas it is stronger the effect of familiarity and trust on labels usage. On the other

hand, comprehension affects both familiarity and trust that are mediators of its effect on label usage. We can, thus, re-read the results of the two experiments in light of this general explanation of label usage and advance the following interpretative hypothesis.

We do not find systematic effects for the nudges tested, not only because they do not activate efficiently cool and reflective cognitive processing, but also because respondents are not so familiar with labels. Low familiarity with labels was also found in the preliminary survey. As familiarity and trust increase with time then an easier and more intuitive processing of the labels might empower their effects. On the other hand, improving the comprehension is also important since we have evidence that comprehension becomes more important as the age of an individual increases. Since car purchase decisions are concentrated in the adult and middle-aged population, working on the comprehension of labels is also important. Hence, the most general policy recommendations is that the Commission and Member States should use the most effective and easiest to understand label in order to gradually improve comprehension and, thus, affect familiarity and trust and eventually impact label usage.

Absolutely exploit the promotional material channel

In order to reinforce and speed up the process of familiarisation, it is of the uttermost importance that the promotional material channel is exploited for conveying to consumers the correct nudges. Promotional material is more familiar to consumers, is less constrained by a fixed structure, and is shown to have some remarkable effects on some of the response variables used.

Labels should be based on:

- Vertical layout;
- Absolute classification system;
- Running cost per mile/km to be rendered with a graphic colour based rating that is placed at the side of the actual amount.

Promotional material should be based on: the approach summarised in the figure below:

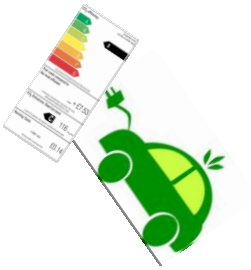
- Graphic element for car classification using absolute classification;
- An additional textual explanation on CO2 emissions;
- Running cost salience (possibly to be rendered with coloured based rating as proposed for labels) together with a footnote explaining the unit of measurement.

TABLE OF CONTENTS

LIST OF EVIDENCE BOXES

LIST OF TABLES

LIST OF FIGURES



Introduction

London School of Economics and Political Science (LSE), Block de Ideas, Universitat Oberta de Catalunya (UOC), and Tech4i2 Limited were commissioned the Study "Testing CO2/Car Labelling Options And Consumer Information" as part of the Framework Contract (No EAHC / 2011 / CP / 01/LSE) on Behavioural Studies. The overall aim of the study was to test in experimental settings the effectiveness of possible new variants of car eco-labels and promotional material. In order to accomplish this overall goal the study included the following key tasks:

- **Task 1: Preparatory Phase**. This comprised: a) a review of the literature; and b) the conducting of a preliminary survey in 10 countries (BE=Belgium;DE= Germany; FR= France; IT= Italy; NL= Netherlands; PL= Poland; RO= Romania; ES= Spain; SE= Sweden; UK= United Kingdom) with a total sample of N= 8000 respondents (800 per country);
- **Task 2: Laboratory experiment testing labels**. A laboratory experiment was conducted at the LSE Behavioural Lab (N=405 respondents);
- **Task 3**: **Online experiment re-testing labels and testing promotional material.**This was realised conducting an online experiment in 10 countries (BE=Belgium;DE= Germany; FR= France; IT= Italy; NL= Netherlands; PL= Poland; RO= Romania; ES= Spain; SE= Sweden; UK= United Kingdom) with a total sample of N= 8000 respondents (800 per country);
- **Task 4**: **Analysis and Recommendations**. We performed econometric and multivariate statistical analysis on the data from the three empirical components listed above and we triangulated the findings to extract policy implications and recommendations.

This is the Final Report of this study where we present the key findings and implications mostly in narrative fashion and with a minimal level of technicalities that are instead illustrated in full detail and depth in the Technical Compendium (available as a separate document presented as Annex II). With respect to the current state of the art this study represents a one of a kind contribution in terms of both scope and experimental approach. We found no study that uses a behavioural experimental approach to test the effectiveness of cars eco-labels with the same extensive scope (first a laboratory experiment in one country and then an online experiment in ten countries, both of which

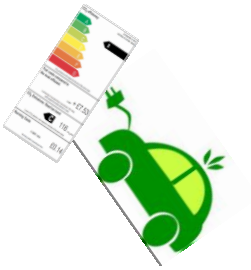
preceded by a survey also in ten countries) and with the same level of sophistication in the randomisation design and procedures as ours. The three empirical components have produced a massive amount of new empirical evidence that we have processed using complex econometric and statistical multivariate analysis techniques. This entails, thus, a very high degree of technical sophistication and complexity that could not be condensed in a report such as this one, which is aimed at a wider policy and generalist readership. We decided, thus, to concentrate all the technicalities in the annexed Technical Compendium (see below) and also to transparently provide to the interested readers all other technical elements in another seven annexes that we list below:

- **Annex I**: **Clean datasets**:
  - Annex Ia Online Preliminary Survey Database
  - Annex Ib(i) Laboratory Experiment Database round one
  - Annex Ib(ii) Laboratory Experiment Database round two
  - Annex Ic(i) Online Experiment Database
  - Annex Ic(ii) Online Experiment Process Tracing Labels
  - Annex Ic(iii) Online Experiment Process Tracing Promotional Material
- **Annex II**: **Technical Compendium**. Here the reader will find full details about:
  - **Literature review**. The extensive review of the literature that subsequently shaped design and methodological choices;
  - **Design**. All the details about the experimental conditions tested (with graphic exemplifications) and about randomisation and experimental procedures and protocols;
  - **Response variables**. Detailed illustration and justification of the response variables used to measure the effectiveness of the experimental conditions tested;
  - **Samples**. Detail about the samples used and related descriptive summary statistics;
  - **Analysis performed**. Technicalities on the econometric and multivariate statistical analyses performed on the data gathered;
  - **Regression and multivariate statistics analyses results tables**. As in this Final Report we present the main results mainly in narrative fashion, all the technical tables and graphs supporting such account are contained in the Technical Compendium;
  - **Car database illustration**. The illustration of the car database we built ad hoc for this study (see below Annex VIII) and how we used it to produce the visual stimuli (see below Annex IV) through which we administered the treatments to the subjects of the experiments
- **Annex III**: **Preliminary Survey tables and graphs**. All the descriptive tables and graphs derived from the processing of the survey questionnaire are reported in this annex;

- **Annex IV**: **Visual stimuli**. The treatments and control conditions for the laboratory and online experiments have been rendered into many graphic images of labels and promotional materials, an exemplificative selection of which can be found in this annex;
- **Annex V**: **Preliminary Survey Questionnaire**. This is the questionnaire used for the preliminary survey;
- **Annex VI**: **Laboratory experiment tasks and questionnaires**. This contains the set of pre and post treatment questions and illustration/exemplification of the tasks of the laboratory experiment;
- **Annex VII**: **Online experiment tasks and questionnaires**. This contains the set of pre and post treatment questions and illustration/exemplification of the tasks of the online experiment;
- **Annex VIII**: **Cars Database**. In this database relevant information for 470 different car models, in ten different countries, have been gathered and then used to produce the visual stimuli.

Therefore, in this Final Report we will only make very brief reference to design, type of analysis performed, and we do it in the most intuitive and accessible way. We do not report technical tables with the results of regression and multivariate statistics analysis but we comment these results in plain language (the tables and graphs can be consulted in the Technical Compendium) In chapter one we present the experimental objectives of the study (§ 1.1), we explain the rationale and the meaning of testing policy options in the domain of car eco-labels with a behavioural approach (§ 1.2), and we very plainly and intuitively illustrate the logic of randomised control trials and the type of analysis done on the data gathered. The following three chapters (chaps 2, 3, and 4) deal each with one of the three empirical components and are equally structured with: a) a brief illustration of key design elements; b) a descriptive, plain, and narrative account of the main findings; c) a recapitulation of such findings (in table formats only for the laboratory and online experiments). In chapter 5 we consider the implications of the three empirical components, first separately and then we triangulate them to provide some general policy implications and recommendations. Finally, for ease of reference to the readers who may not be interested in all the technicalities of the Technical Compendium but may want to have a look at the graphic rendering of the tested experimental conditions, in the Appendix to this report we list the key tested experimental conditions with their graphic illustration. In this respect we must stress that, as a function of randomisation, we have produced and shown hundreds of different visual stimuli (between the laboratory and the online experiments) of which both the few reported in this Appendix and the more numerous ones contained in Annex IV are only a limited exemplificative selection.[2]

[2] The reader interested in this particular aspect may look at chapter 6 of the Technical Compendium (Annex II) to understand the logic by which visual stimuli were produced from the information contained in the car database and simultaneously assigned randomly to the subjects of the experiments.



1 Overall approach

1.1 Context and objectives

The Directive 1999/94/EC ('CO2 Car labelling directive') prescribes some general aspects for the format of car eco-labels (CO_2 emissions to be displayed near each passenger car at the point of sale) and requires that promotional material contains fuel consumption and specific CO_2 emissions data of the car models to which it refers. Within these broadly defined prescriptions a number of different labels are currently used in the different Member States that have adopted them. Two recent studies, one commissioned by the European Parliament(Grunig et al., 2010) and one by DG CLIMA of the European Commission(Branningan et al., 2011), show that there is wide variability in the way Member States have implemented the directive. These two studies also call for the gathering of more evidence on the effect that the directive has had on consumers' awareness and purchasing behaviour and they explicitly suggest the need to apply a behavioural approach. The variability and, at times, the lack of clarity (especially for promotional material) in the formats adopted in different Member States can reduce the effectiveness of labels and promotional material[3]. This creates a need for the Commission to intervene with proposals to standardise and make labels and promotional material more effective. Against this background with this study DG CLIMA aimed to test:

- **The effectiveness of car labels**. Test the effectiveness of alternative car labels, with specific reference to their content and layout, in terms of informing potential car buyers about, and raising awareness of, CO2 emissions, fuel efficiency, and running costs of new cars;
- **The effectiveness of mandatory information in promotional material**. Test the effectiveness of mandatory information in promotional material in terms of informing

[3] It is worth noting that variability in the form and quality of labels and promotional material undermine the relevance and usefulness of environmental information (Yates, 2009). As for the text of promotional material one should also bear in mind that consumers tend to think that 'if it is hard to read then it is hard to do' (Song & Schwarz, 2008). In other words unclear or generic promotional material may end up be ignored and have no effect whatsoever.

potential car buyers about - and raising awareness of - CO2 emissions, and fuel efficiency of vehicles;

As we illustrate in greater detail in the relevant chapters (chapters 3 and 4) and in the Appendix (chapter 6), after exploring the car purchase process in the preliminary survey (chapter 2), we pursued these objectives by testing in an experimental settings (first in the laboratory experiment and then in the online experiment) several types of information that are already contained in existing labels and promotional material or that could be included in the future. We tested them basically through simulated car purchases (experimental tasks) that the participants had to perform in the experiments. We, thus, registered their choices (behavioural measures). In addition we elicited cognitive effects ("noticeability"; "recall and processing of information") asking questions to respondents after they were exposed to the experimental conditions (i.e. post-treatment questionnaire). These questions were opportunely phrased to construct psychometric scales (self-reported measures). In simplified fashion we can say that we tested the effectiveness of two basic groups of information elements:

- **Standard elements**. We mean elements that, though in different formats, are already contained in most of the labels and promotional materials currently in use, such as for instance the different CO2 classification systems (absolute: car rated compared to cars from all classes ; relative: car rated compared to cars belonging to the same class; combined: a combination of absolute and relative rating) and their format (text only, in graphic format using either a vertical or a horizontal layout, see Figure 19, p. 99);
- **Additional elements**. We mean elements that either are not included in currently used labels and promotional material or are included only in few countries. Elements such as :
  - Information about running costs[4];
  - Information about taxation related to CO2 emissions;
  - Information about the lost fuel savings that buying car X implies if compared to the most efficient car Y in the same class;

These experimental conditions were administered to participants by way of showing them visual stimuli containing different combinations of the information elements. These visual stimuli can be found both in Annex IV and in a more restricted number in the Appendix of this report (chapter 6). They have been defined, also in view of the results of the preliminary survey, in agreement with DG CLIMA and following a Stop & Watch

[4] We use this general expression to refer to the running costs in terms of fuel only and we tested it as such without adding the other elements of maintenance costs (i.e. tires, oil, other consumables, insurance, etc.). We have agreed this choice with DG CLIMA mainly on the basis of two reasons, one substantive and one pragmatic. First, the piloting of the experiment showed that most people intuitively associated running costs to fuel. Second, for the online experiment in 10 countries it was beyond feasibility to find comparable and consensual measures of the other maintenance cost elements.

approach in which at the end of each empirical step the results were presented and discussed with DG CLIMA before moving to the next one.

It is important to anticipate here a relevant aspect that is later further explained in § 3.1 (in slightly more technical fashion). This aspect is key to both the understanding of the terminology used and especially to the correct interpretation of the results.

In our exposition we use at times the generic expression visual stimuli or visual stimulus in place of labels or promotional material. The visual stimuli are the channel through which we administered to our participants the experimental conditions tested: the individual pieces of information that we illustrated above. We test, in fact, these pieces of information we do not test labels and promotional material as such.

A label (or a promotional material) is a set of pieces of information and their graphical presentation. For example, we can include the absolute emissions' classification in a vertical format, together with the information on running costs expressed as euro per km. As one can imagine, there are various formats for both types of information (absolute, combined, relative, graphical, textual, vertical or horizontal for CO2 and per mile, per month, per year etc. for the running costs). In principle it might be possible that, just to give an example, vertical absolute has a different effect when combined with running cost per month as compared to when combined with running costs per mile. As a result, there are many possible 'interaction effects' that one may possibly want to consider if the objective is to test different labels as the treatments. In the technical jargon in order to capture such interaction effects a 'full factorial' design is required, whereas we could only perform a 'main effect' design[5]. A full factorial design was beyond feasibility for this study given the very large number of information elements that we were asked to test.

Our aim and subsequent claims, thus, are more restricted: we tested the effect of a single piece of information (or a specific format) at a time. Our empirical strategy is e.g. to detect if the inclusion of running costs per km changes the response by the subject (better recall, higher likelihood to buy a more eco-friendly car, more cognitive processing), regardless of the other information that is included on the label. This means that, we repeat, we did not test specific labels or promotional material as such, which is to say as holistic treatments. We tested the information elements as shown in different

[5] In the laboratory experiment the limit was sample size, whereas in the online experiment (where sample is large) the limit was that 20-25 minutes is considered as the maximum duration compatible with obtaining reliable data (after due to fatigue and to the fact respondents are not in a controlled environment as in the laboratory, they may simply provide top of mind answers or perform experimental tasks just to finish them).

visual stimuli. So, if we state, for instance, that 'running costs per 5 years' are effective with regard to a given variable, this means only that this information element, as contained in a number of visual stimuli, is effective. As a result our findings support or reject the use of single pieces of information (and their format) but cannot be used to infer the 'optimal' (e.g. the most effective) label, because the latter imply testing the effect of showing all the possible combinations of the individual pieces of information (and their format), which was beyond the possibility of this study.

1.2 Behavioural foundations for consumers' policies

Eco-behaviour in car purchasing decisions concerns **"Credence Attributes"**(Teisl et al., 2008, pp. 143-144), which means elements that cannot be as easily verified by consumers as elements directly visible or experienced such as for instance colour, road handling, acceleration, etc. The colour of a car or its acceleration power, not only can be directly experienced, but also directly tap into consumers' emotions and intuitive thinking, a fact that manufacturer's brand campaigns saliently leverage. Eco-friendliness and fuel efficiency are indeed related and could be objectively verified, but the related information is either provided in unclear and unfriendly way or is simply not processed by consumers due to what behavioural scientists call heuristics and biases (see below). This means that: a) choices are not always optimal both for individual consumers and for society as a whole; b) simple provision of additional information does not have a linear impact and is mediated by perceptions, attitudes, experiences, and other personal characteristics. The review of the literature presented in the Technical Compendium (chapter 2) fully confirms these points and others such as: the presence of an 'attitudes-action gap' (consumers are aware about environmental impact of cars but this does not translate into eco-friendly car purchases) and inconsistent preferences (self-reported importance of fuel economy and then purchases of fuel inefficient models), a widespread level of eco-ignorance and lack of understanding of the connection between CO2 emissions and fuel efficiency. As noted in a report by the UK Department of Transport commenting survey results: ''*Concern for the environment in general and the environmental impact of cars, which is evident, does not often translate into behavioural change at an individual level*''(2004, p. 7). There is also clear evidence that standard consumer information policies cannot offset consumers' biases and that providing more information in certain circumstances helps manufacturers confound consumers (Heinzle & Wüstenhagen, 2009; Waide, 2001). It is worth anticipating that these insights extracted from the review of the literature have been mostly confirmed ex post by the results of our preliminary survey in 10 European countries.

The perspectives of what is more widely known as behavioural economics[6] are highly relevant to explain the inconsistencies in preference, eco-ignorance, and an attitude-action gap that characterise the understanding of eco-behaviour in car purchase and to help policy makers devise measures contributing to better choices both for the environment (CO2 emissions) and for the budget of consumers (fuel efficiency). We try to summarise below a huge body of literature in an extremely succinct and simple way as to then illustrate the rationale for the options that we tested and for the way we tested them.

Behavioural Economics (henceforth also simply BE) has effectively challenged with both theoretically and laboratory controlled experiments the standard view of the rational agent processing all information and making the more optimal self-interested choices. It has shown that: a) this model is unrealistic, since in many situation choices are made with little thought even when information is available (so information is not processed); b) in the brain processing of the stimulus embedded information, emotion and cognition, always interact and shape motivation. Starting from these premises behavioural economists have presented evidence-based critiques of the standard economics view of human behaviour and demonstrated the human limits of computational power or will power. In brief, this has been vividly and accessibly pictured by Kahneman (2011) with the distinction between System 1 and System 2 as the two components of our minds that always interact in processing information and in making decisions / taking actions. The main characteristics of System 1 (reflexive) and System 2 (reflective) are summarised in the next table.

System 1 is automatic and affective (highly impacted by emotions) and effortlessly generates impressions and feelings that, when not challenged through a more reflective consideration, are the main sources of the explicit beliefs and deliberate choices of System 2. System 1 relies on mental "shortcuts"; it quickly proposes intuitive answers to problems as they arise. These basic assessments play an important role in intuitive judgment, because they are easily substituted for more difficult questions—this is the essential idea of the heuristics. Such heuristics, however, can be the sources of biases in analysing situations and responding to stimuli and can lead to systematic mistakes.

[6] In reality the work that today goes under this newly fashionable disciplinary label include contribution coming mostly from cognitive psychology, social psychology, and to a lesser extent from what one could be better term as 'experimental economics'. The broader label 'behavioural science' would therefore seem more appropriate. We decided, however, to stick to the more popular 'behavioural economics' as not to enter into irrelevant disciplinary dispute and use an expression that may be more familiar to a generalist and non specialist readership.

Such biases may be harmful to consumers and, as a result of cumulative choices by many consumers, can eventually become harmful also to society as a whole. On the contrary, System 2 follows controlled processes, is slow, effortful, conscious, rule-based and can be also employed to monitor the quality of the answer provided by System 1. System 2, however, is constrained by human capacity or will to always process all the available information to make optimal choices.

Table 1 System 1 and System 2

System 1 (fast)	System 2 (slow)
Quick, automatic, no effort, no sense of voluntary control. Continuous construal of what is going on at any instant	*Slow, effortful, attention to mental activities requiring it. Good at cost/benefit analysis, but lazy and saddled by decision paralysis (cognitive overload)*
Characteristics • Quick (Reflexive); • Heuristic based; • Use shortcuts;	**Characteristics** • Deliberate (Reflective); • Conscious; • Rule-based;
When it comes into play • When speed is critical; • Avoid decision paralysis; • When System 2 is lazy or not activated (not worth, no energy, lack of awareness)	**When it comes into play** • To take over when System 1 cannot process data; • To correct/override System 1 if effort shows that intuition or impulse is wrong;

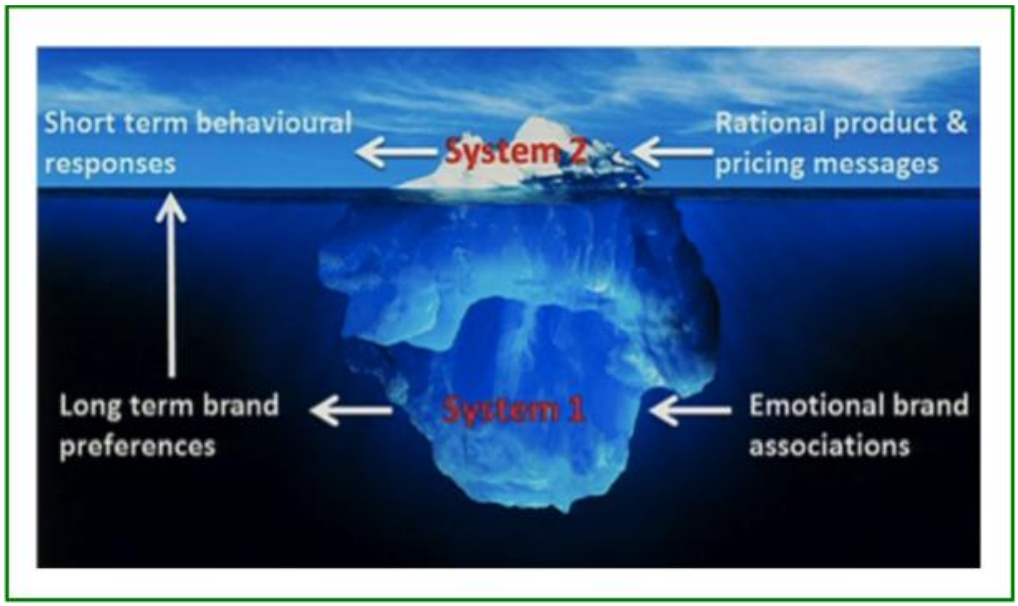
Source: our elaboration from Kahneman(2011) and (Thaler & Sunstein, 2008)

System 2 may remain in stand-by if individuals are not triggered (usually in emotional ways) or it may stumble into 'decision paralysis' (cognitive overload) and thus let System 1 rule our choices and actions. There are a number of biases identified by BE that are relevant in our domain of interest, either as factor explaining eco-indifferent behaviour or as factor that may be exploited to induce more eco-friendly purchasing behaviour; they are discussed at some length in the Technical Compendium (paragraph 2.2). For instance, since individuals show consistent loss aversion (losses are negatively felt much more than how gains are enjoyed), framing eco-labels information in term of losses rather than gain may have an effect on purchasing decisions. It has been shown that consumers tend to think that 'if it is hard to read then it is hard to do' (Song & Schwarz, 2008). In other words unclear or generic information may lead to 'rational ignorance': the way in which the information is provided may wrongly lead individuals to conclude that the cost of

educating oneself on an issue exceeds the potential benefit that the knowledge would provide.

The fact that System 2 is only the tip of the iceberg and System 1 is the submerged and much greater part has been recognised and used in marketing much before policy makers turned their interest to the behavioural foundations of policy. This is impressionistically conveyed in the figure below where we included an image taken from an online marketing magazine. Brand campaigns in many industries and especially those of cars manufacturers tap directly into emotions and consider pricing and rational product messages of secondary importance.

Figure 1 A marketing perspective on System 1

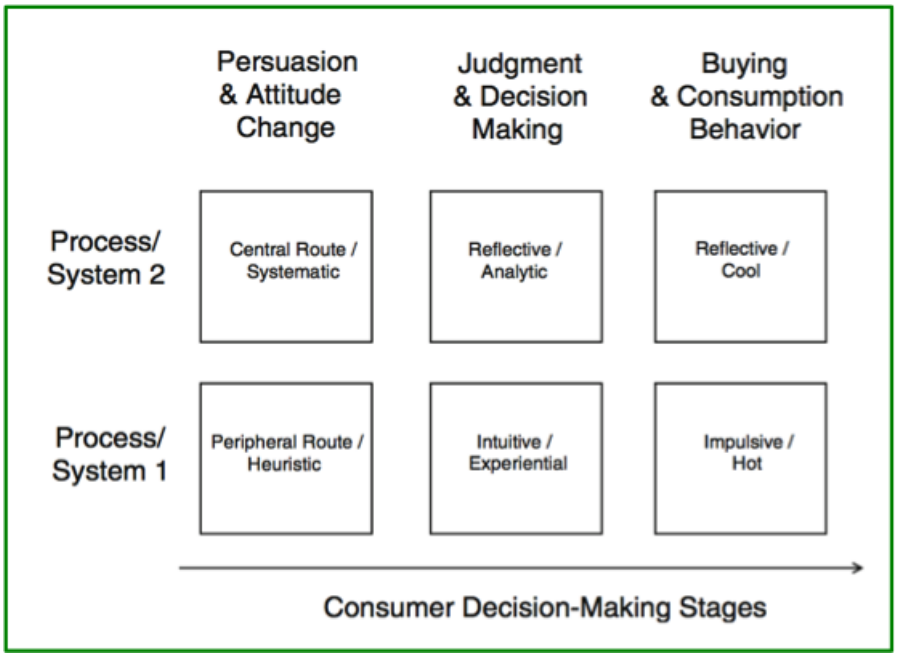


Source: http://bdecastella.com/2012/12/effectiveness-the-long-and-short-of-it/

The best seller book "Nudge"(Thaler & Sunstein, 2008) represents the policy manifesto of how behavioural insights can help policy makers use the same instruments to counter-balance industry strategies and help consumers make better choices for themselves and for their community without reducing their freedom. They called these instruments 'nudges' and argue that they amount to designing 'choice architectures' that leave freedom of choices but can be seen as 'de-biasing' mechanisms. The main thrust of the 'nudge philosophy' is that to influence and eventually change behaviour, policymakers stand a better chance by reducing the restraining factors rather than by increasing the driving factors. Nudges should make it easier and to some extent effortless for consumers to make more sensible choices. It is a somewhat counter-intuitive approach as the standard tools most of us use when trying to change others' behaviour are arguments,

promises, and threats. It is less common to look for ways of making it easier for the other person to do the right thing. Yet, starting from the assumption of limited human power to rationally process information, nudges aim to frame choices in such a way that either consumers have to make minimal effort or they are forced to activate reflective cognitive processes. The quintessential and most effortless nudges concern defaults, since for reasons of laziness or distraction (status quo bias) many people will take whatever option requires the least effort, or the path of least resistance. This implies that if, for a given choice, there is a default option, the latter is what will be picked up in most of the cases by the decision-maker. Fixing as the default option the optimal choice (from the point of view of society and/or the policymaker) can produce policy desirable results. An alternative is to provide stimuli that may activate unconscious drivers leading to desired choices by tapping directly into emotions. In general, looking at the figure below, which provides a more granular view of the distinction between System 1 and System 2 across the various stages of consumers behaviour, when defaults or direct emotions tapping nudges are not applicable policymakers should devise and frame nudges that stop 'hot cognition' and activate 'cool cognition'.

Figure 2 Stop hot and activate cool cognition



Source: Strack, Werth& Deutsch (2006).

One of such possibilities is, for instance, 'mental accounting' (Thaler 1985) whereby consumers are pushed to construct and consider a budget in relation to a particular purchase decision. Another one is, as anticipated, framing information nudges as 'losses'

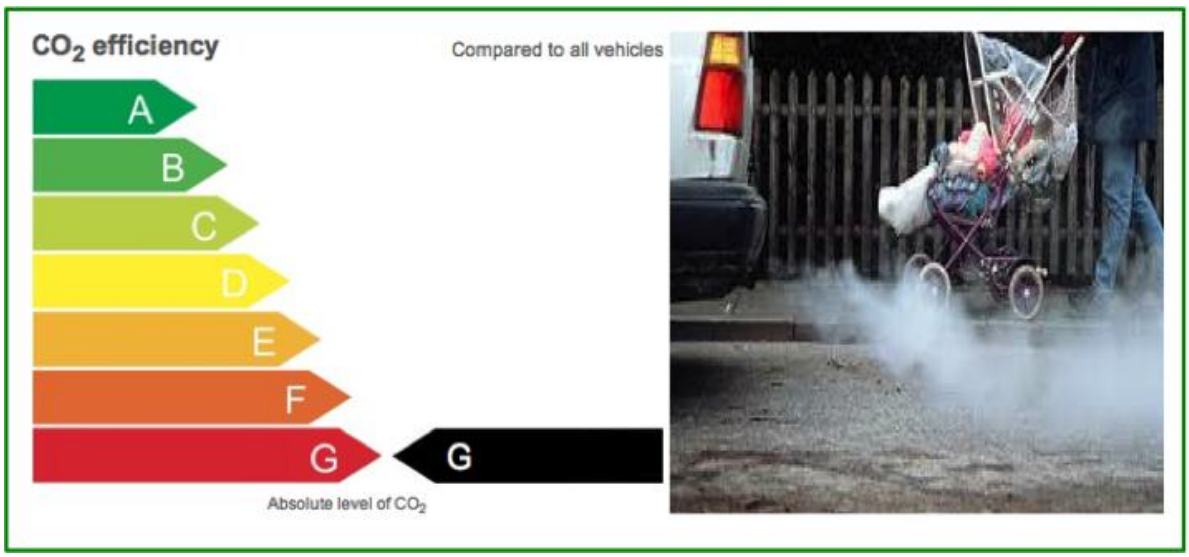
(if appropriate). A third one is providing information in clear and large font or graphically to avoid 'rational ignorance'.

How do the considerations in this paragraph bear on our study? In the several ways we explain in the next paragraph, but we can anticipate with the following questions: what variants of labels and promotional materials we tested and why? How we measured their effectiveness? How we tested them and recovered their true casual effect?

1.3 Key aspect of design and analysis

The first way in which the insights of the behavioural approach discussed above shaped our study concerns the nudges (treatments) we decided to test given some constraints. The variants of information and layout elements that we tested for labels and promotional material were agreed with DG CLIMA. The agreed choices were to some extent constrained in the sense that they could not deviate too much from general regulatory prescriptions and from the labels and promotional material that are already currently in use. For car labels there was not any possibility for default type nudges, and neither for emotions arousing nudges, such as the hypothetical one reported in the figure below, for they fall outside the framework set by the earlier mentioned directive and by labels and promotional material currently in use.

Figure 3 Fictitious car eco-label example



The variants of labels and promotional material that we tested, thus, fall into the category of nudges that, by better framing of the existing information elements and/or by including clear and easy to understand new elements, aim to activate reflective and cool cognition in consumers. They should in principle instil doubts on - and stop - automatic associations with those pleasurable and hedonistic aspects experienced or induced by brand campaign (speed, power, status symbol, etc.) and nudge consumers to consider

environmental aspects (public good) but also aspects related to the fuel economy beneficial to them as self-interested individuals (individual financial utility). We did so mainly in two ways; first, we framed better, in terms of presentation (graphic layout and/or more salience for textual information), the information elements contained in our variants as compared to those of labels and promotional material currently in use (used as control conditions). Second, by adding elements, opportunely framed (i.e. lost savings in fuel aimed at exploiting the loss aversion bias) and saliently presented, which are not contained in existing labels and promotional material. For instance the reader can compare, in terms of clarity of common elements and of additional elements, the standard UK label (Figure 18, page 99) used as a control condition in the laboratory experiment, with all other labels used in that same experiment to convey the various nudges tested (Figures 47-51, pp. 100-103); or the greater salience of the information on CO2 emissions contained in our variants of promotional material (Figure 57-59 pp.107-110) compared to the standard version in use (Figure 56 p. 106).

The second way in which the behavioural approach is relevant to our study concerns the variables we used to measure the effectiveness of the various 'nudges' (henceforth we refer to these as 'response variables') and leads us to explain here the distinction between 'behavioural' and 'self-reported' response variables so that we will not have to come back to this issue in the remainder of this report. Self-reported response variables refer to those generated through the answers, that participants to the experiments provide, to a set of questions we asked after they completed the core experimental task (see *infra*). In other words, participant self-reported information is the outcome of the measured variable. Behavioural response variables on the contrary simply amount to the actual choice made by participants and automatically recorded by the experiment programming software. In our case we ask participants to perform a simulated car purchase (see Figure 4, p. 32) and our software automatically record which car they selected (its price, class, model, etc.) in combination with the particular nudge the subject has been exposed to. A behavioural approach would prefer behavioural response variables to self-reported ones, due to the fact that the latter may be affected by the very same biases that we discuss above. For instance, imagine the case in which a label is effective in convincing consumers to pay a higher price for a more eco-friendly car but then the respondents show difficulty in recalling the information that the same label contained. How would this be possible? The answer lies in the fact that a label might affect a user unconsciously[7] and therefore the self-reported measurement does not capture such effect. Conversely, there can be clear gaps between self-reported preferences or intentions and actual behaviours either because answers are affected by well known social desirability effects (answering to please the researcher or in a way

[7] The unconscious here referred is what is intended in cognitive science as any computational operations that the human minds do of which we are aware only of its outcomes.

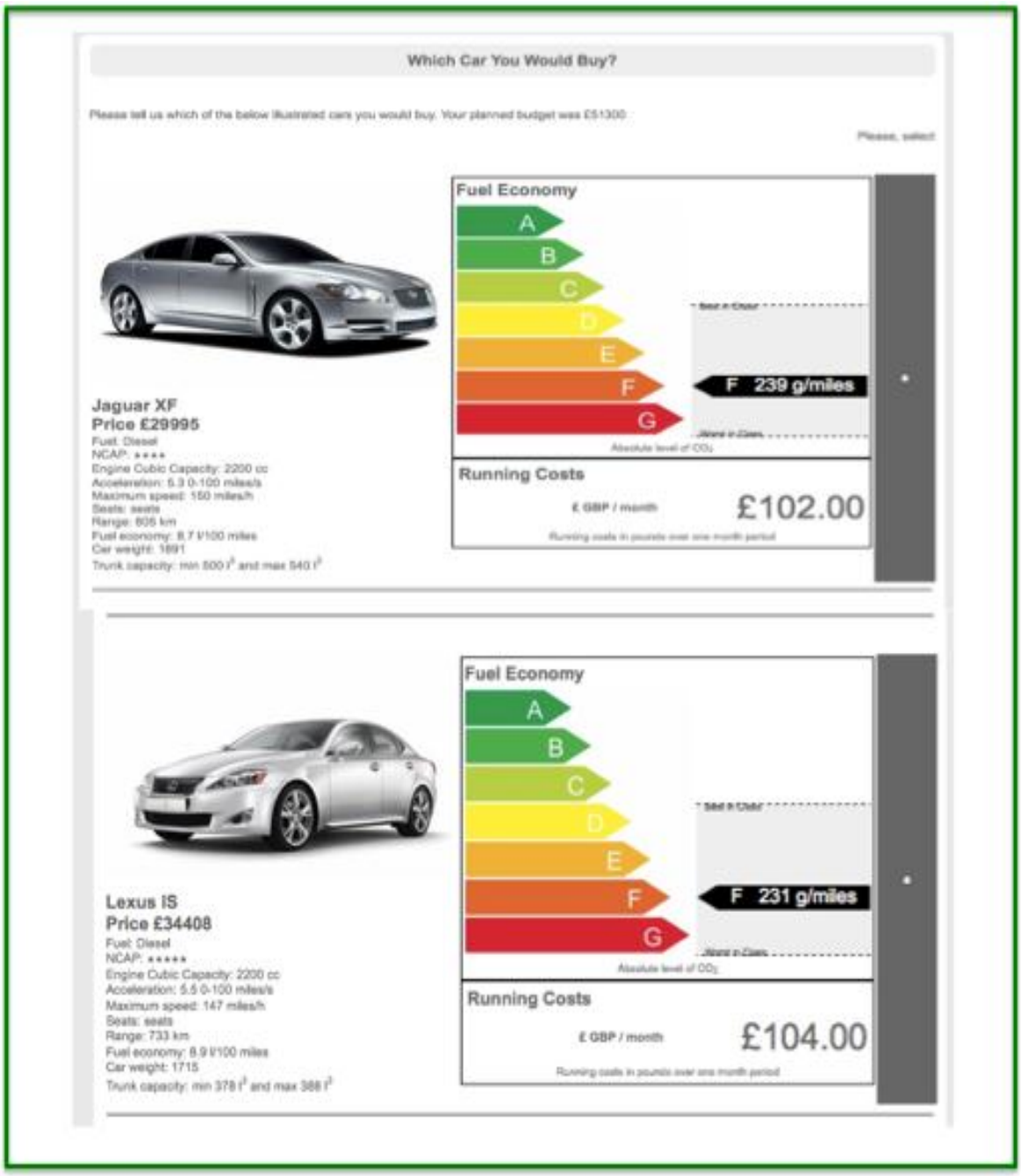
considered generally acceptable) or because self-reported intentions conflict with inner emotional factors.

The core of our two experiments were simulated car purchases to retrieve as a behavioural response variables the Willingness to Pay (WTP) and assess the extent to which the various information elements tested in respectively the labels or in the promotional material had a positive effect in either leading consumers to spend more for eco-friendly and/or fuel efficient cars or in spending less for eco unfriendly and/or fuel inefficient ones. In the Laboratory experiment (with complex randomisation procedures explained in chapter 4 of the Technical Compendium) we presented the participants with 3 or 4 cars associated with the visual stimuli (labels) containing the nudges and asked them to buy one (see Figure 4, p. 32). They were also provided with the price of the car and reminded of the budget they had declared they considered for their next car purchase in the pre-treatment questionnaire (although they can violate their initial plan, for reasons explained below). In the online experiment (both for nudges contained in labels and promotional material we used instead a Multiple Price List (MPL) instrument (Figure 5 p. 33): we showed a car in combination with a label or a promotional material and asked them to choose at what price they would buy it from among the list of prices presented.

When respondents completed the experimental tasks, both in the laboratory experiment and in the online experiment, they answered the set of questions reported in Figure 6 (p. 34).[8] The answers by the respondents were used to construct cognitive self-reported variables of two types: "noticeability", and "comprehension and recall". The questions ask to express an evaluation on several parameters of the car they had seen (eco-friendliness, fuel efficiency, etc.). We used the answers to measure how much they noticed the information and whether they had comprehended and were capable correctly recalling it. In the first case we simply measured whether respondents who were exposed to the nudges noticed eco-friendliness, fuel-efficiency, etc., more than respondents that were exposed to the control condition. In the second case we compared the evaluation expressed by the respondents with the objective information available in our car database as to verify whether or not they recalled the information correctly in general and more correctly than respondents exposed to the control conditions. The exact procedure by which such kinds of variables were constructed is reported in full in the Technical Compendium (§ 4.2). So, basically we have used as measures of effectiveness three sets of variables (WTP behavioural; noticeability and recall cognitive self-reported) plus some variations explained also in the Technical Compendium (§ 4.2).

[8] The precise formulation of the questions contained in the figure changed depending on the type of engine (conventional, electric, and hybrid) of the car shown to respondents. The questions in the figure are only a selection of all those asked but render well enough the logic used to measures self-reported cognitive response variables

Figure 4 Example of simulated purchase (based on conventional engine models)[9]



[9] For the sake of graphic resolution we have placed only two cars in the figure above, but in practice in the simulated purchase subjects were shown 3 cars (and in a few cases 4 according to randomisation design). Please note also that the first car has higher emissions but lower running costs compared to the second due to the fact that it consumes less fuel (8.7 litres per 100 miles versus 8.9 litres per 100 miles).

Figure 5 Example of Multiple Price List used

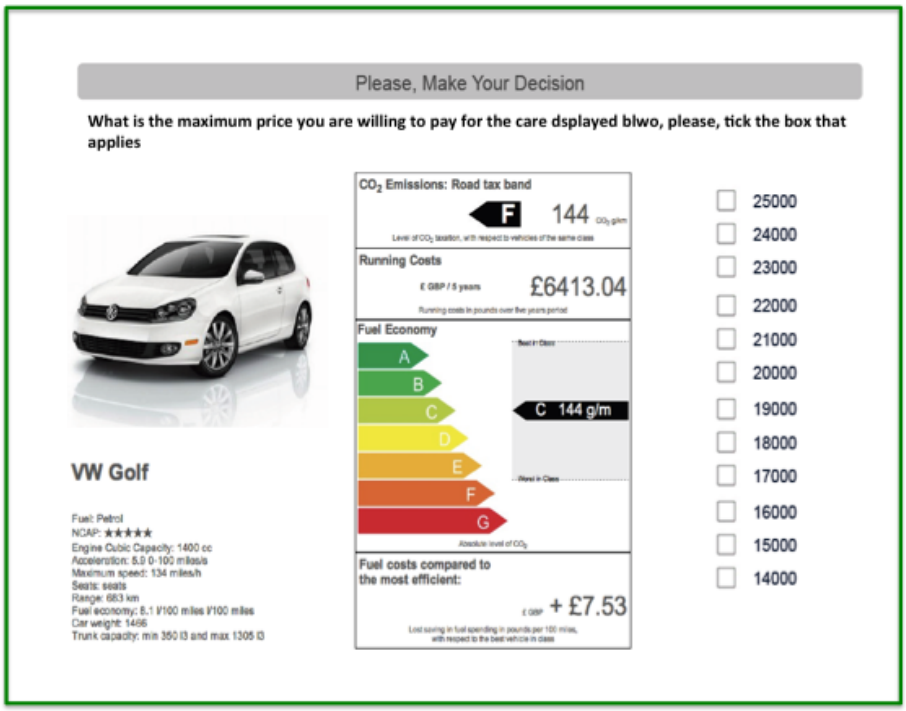
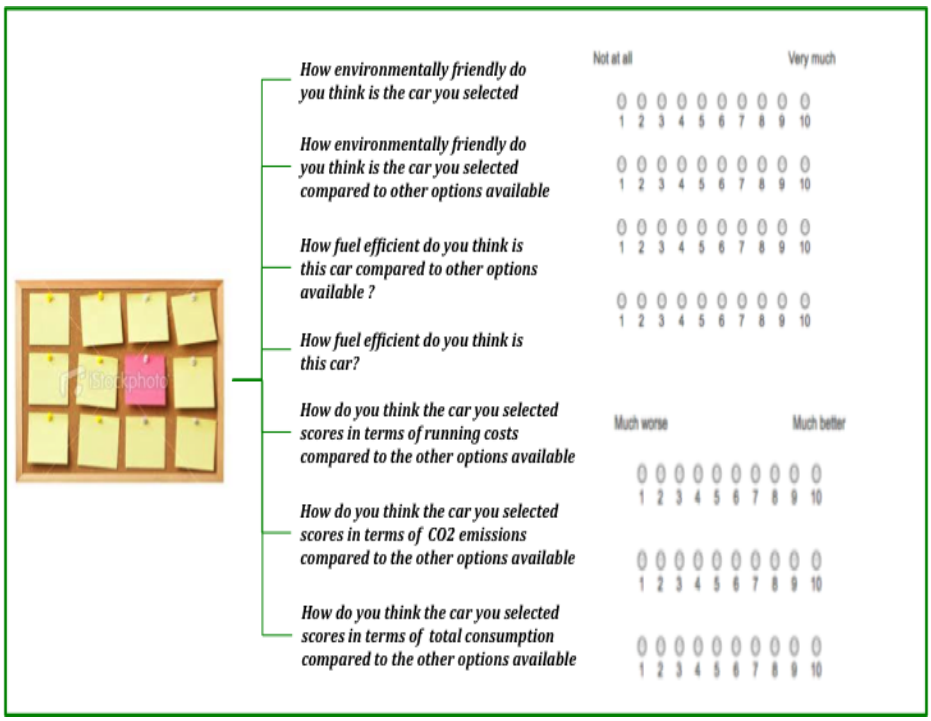


Figure 6 Cognitive measures (exemplificative)



Additionally, in the online experiment both for labels and promotional materials, we used two behavioural measures that elicited the respondents' choice of preferred information. First, for a randomly selected sub-sample we used the 'process tracing' technique illustrated in the figure below and consisting in showing to respondents a blurred picture of the visual stimuli and car for the MPL task. In advance to this, the respondents are provided with instructions explaining them to mouse over the items they are interested to see fully in order to complete the task. When they do so, the respondents leave 'behavioural' traces of the information they need/prefer, which were automatically captured by our supporting software. Second, another randomly selected sub-sample was explained the information that can be contained in promotional material and then asked to choose in which format they would prefer to see it, so that again we record a behavioural choice of preferred information.

Figure 7 MPL task with process tracing

The last among the questions, anticipated at the end of the previous paragraph, that we need to answer is 'how we tested the variants of labels and promotional material and recovered their true causal effect?

The answer is pretty simple: as robust randomised control trials with double randomisation of participants and treatments (i.e. 'nudges') in a subject design where we

have treated and control groups[10]. The treated respondents are exposed to the variants being tested, and the control group to labels and/or promotional material as they are currently used. The allocation of subjects to one or the other is random, i.e. is not correlated with other unobservable variables. The next two figures graphically illustrate this procedure for the laboratory and the online experiments in just an exemplificative fashion.

Figure 8 Treatment and control for the laboratory experiment (exemplificative)

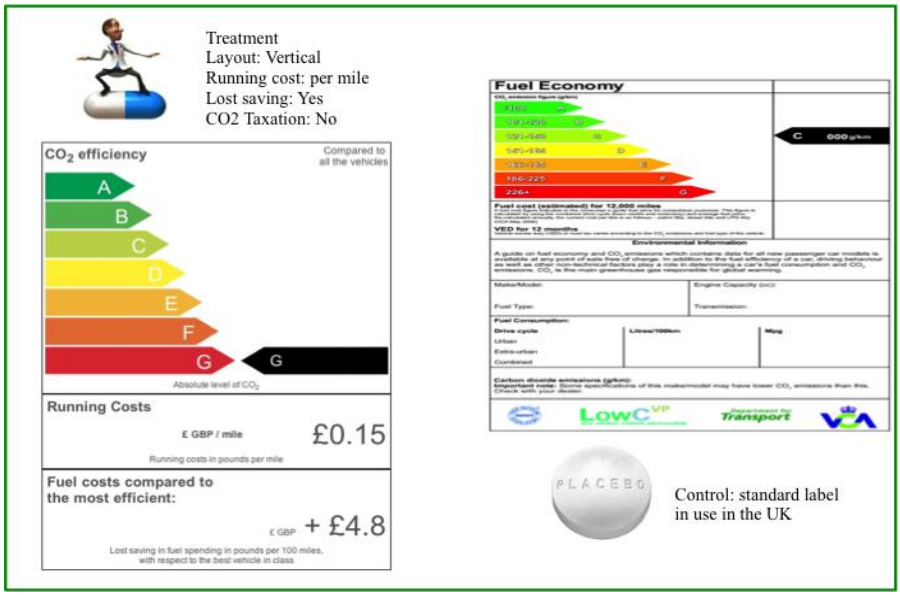
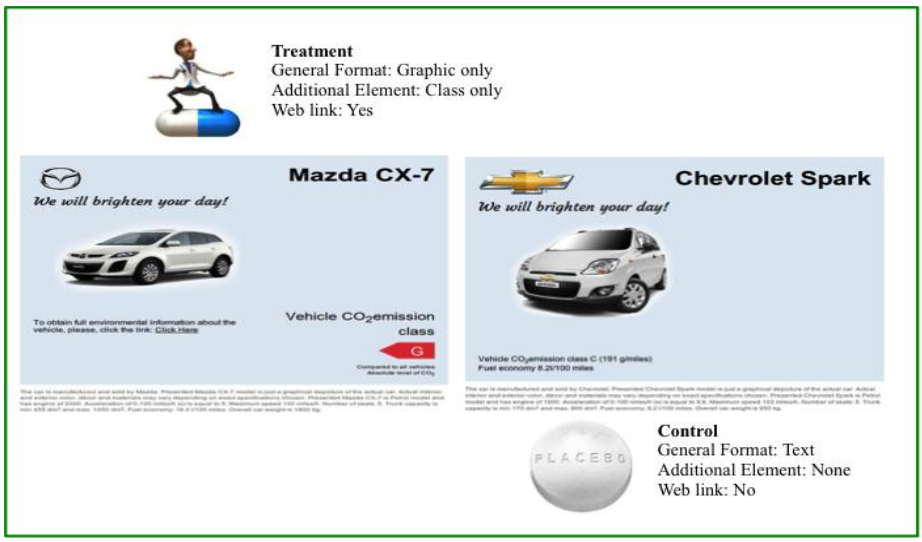


Figure 9 Treatment and control for the online experiment (exemplificative)

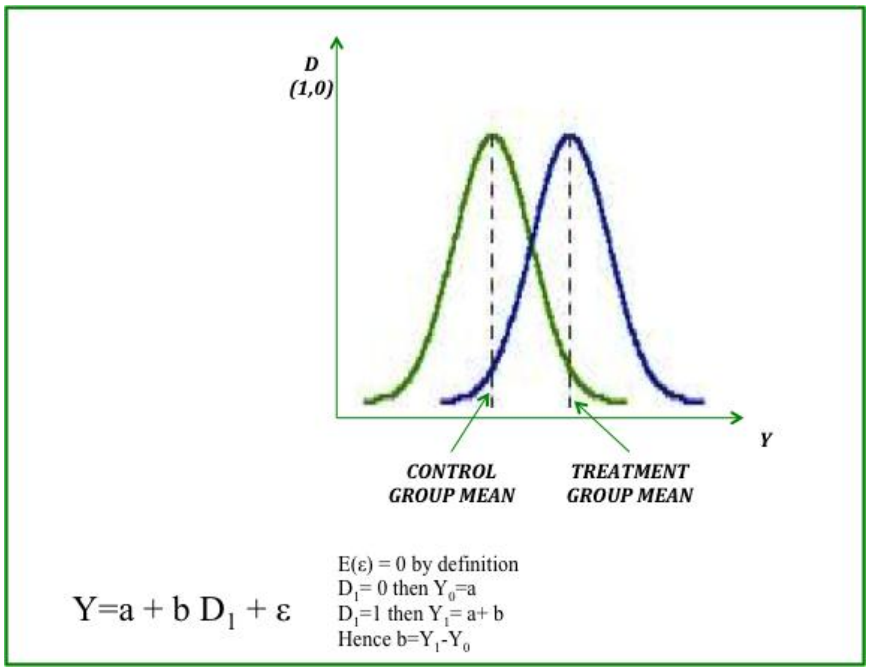


10 As explained in more details in the Technical Compendium (§ 4.1 and § 5.1) in some cases is between-subjects with repeated measures in order to have enough statistical power. This means that respondents are allocated to more than one experimental condition.

In the Laboratory experiment conducted in the UK those subjects randomly allocated to the control conditions saw a number of cars in combination with the label in use in that country and then performed the experimental tasks. Those allocated to the treatment saw instead a number of cars in combination with several possible variants of labels containing the different nudges. In the same way, in the online experiment, subjects were randomly allocated to see cars (and perform the related experimental tasks) in combination with either a control or a treatment both rendered as visual stimuli. In the figure above we have placed the control condition for promotional material and only one example of the variants tested.

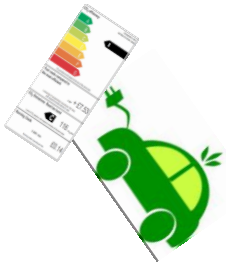
This design guarantees that the difference in mean between the treated group and the control group for any measured response variables is due to the causal effect of the treatment (nudge) on that variable (see figure below). If, for instance (numerical example is fictitious just for the sake of intuitive explanation), participants shown a visual stimulus, with running cost information, express on average a willingness to pay for a certain car in a given class of € 25.000 and participants (control group) who were randomly allocated to a visual stimulus, not containing this piece of information, on average expressed a willingness to pay for the same class of cars of € 20.000, then the € 5.000 difference in mean is the estimation of the causal effect of the running cost nudge on the response variable willingness to pay.

Figure 10 Treatment effect: a difference in means



On this premise we used regression analysis to recover the 'treatment effect' of the tested variants on the response variables used. The interested reader will find a more detailed and technical explanation in the Compendium (§ 4.4 and 5.4) as well as all the technical and graphical outputs of regression analysis and of other multivariate statistics analyses (Compendium, § 3.6, § 4.5, § 5.5).

The last technical element that we need to explain before moving to present the findings in the next chapters concerns statistical significance. The difference in means that we recover through regression analysis is presented in the technical tables of the Compendium, in the form of coefficients. In order to conclude that a nudge X is effective on a response variable Y we would like to have the right sign of this coefficient and next to it also one or more asterisks indicating that the coefficient is statistically significant. In most cases we expect a positive sign meaning - as in the hypothetical example made earlier - that the mean for the treated group is larger than the mean for the control group. If we have the opposite sign, then this would mean that the status quo (control group) is more effective than the proposed variants (nudges or treatments). When we test for the statistical significance of a coefficient we want to know if we can reject the so called 'null hypothesis' that the coefficient is simply obtained by chance rather than being a robust measure of a real causal effect. If the test rejects the null hypothesis we conclude that the coefficient is statistically significant and can be taken as a robust measure of the causal effect of a given treatment in the specified sense that on average it produces on the response variable a larger effect than the control condition. On the other hand, lack of statistical significance does not imply that the control condition is better than the treatment; it is just an inconclusive result.

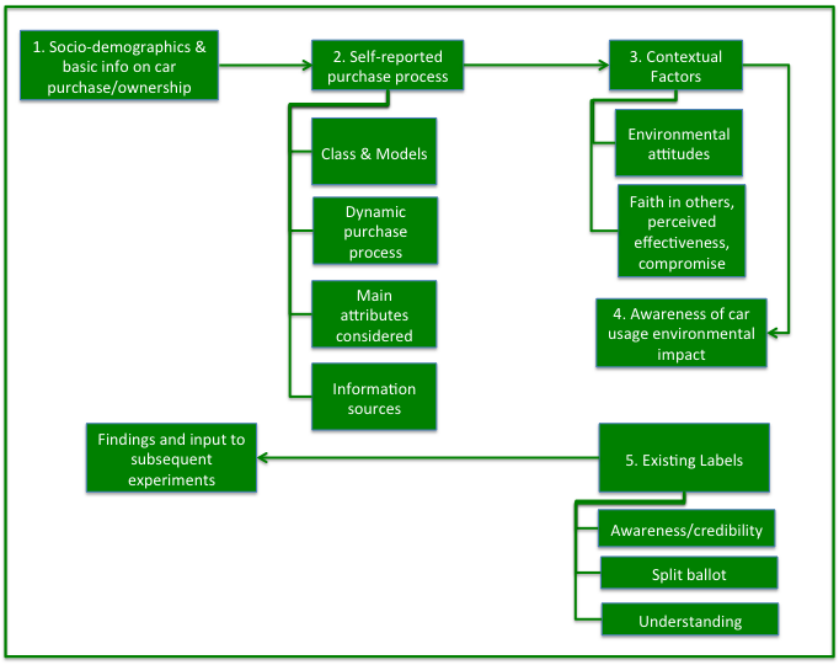


2 Preliminary survey

2.1 Design

The review of the literature providing the rationale for the design of the survey, the full illustration of this design, and the detailed description of the sample are presented in the Technical Compendium (respectively in § 2.1 and in § 3.1 through § 3.4). Additionally the interested reader can consult the full version of the survey questionnaire (**Annex V**). Hence, below we very succinctly recall the objectives of the survey, the overall design of the questionnaire, one peculiar feature (i.e. *the split ballot*), and the key sample parameters. The objectives of the survey were: a)to explore the car-purchase process and the specific (main attributes such as price, safety, fuel efficiency, etc.) and contextual (self-reported attitudes and intentions) parameters shaping it; and b) to investigate the awareness among consumers of the environmental impact of car usage.

Figure 11 Conceptual structure of the survey questionnaire



Following these objectives and coherently with the literature review, the questionnaire has been designed along the block reported in the figure above. The main part of the questionnaire consisted largely of standard survey questions, but toward the end respondents were randomly split into four sub-samples, each allocated to see the same car (a VW Polo) but matched to four different simplified labels containing only the CO2 emissions using four different classification systems (see graphic illustration in Figure 17, p. 99).

5. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the Relative Classification system (i.e. a vehicle is rated compared to vehicles belonging to the same class);
6. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the Absolute Classification system (i.e. a vehicle is rated compared to vehicles from all classes);
7. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the Combined Classification system (i.e. a combination of absolute and relative rating);
8. Participants were shown a variant of the VW Polo associated with a simplified Car Label including the German classification system (a sub-variant of relative classification).

Then they were asked four questions about the car model they had just seen associated with one of the four classification systems to test the extent to which such systems had a noticeability effect:

- Q37a: 'How environmental friendly do you think this car is?'
- Q37b: 'How fuel efficient do you think this car is?'
- Q37c: 'How environmentally-friendly do you think this car is compared to other similar type of cars?'
- Q37d: 'How environmentally friendly do you think this car is compared to the car/s in your household?'

The split ballot was only an exploratory test good enough to shape the design choices of the laboratory experiment but did not aim to provide a conclusive judgement on the relative effectiveness of the four classification systems.

Next table below summarise the key technical parameters used to draw the sample of 8000 respondents in 10 countries that answered the questionnaire from a panel representative of the online population in these countries.

Table 2 Preliminary Survey sample: summary technical parameters

Population	General population aged 18 to 65 years old
Scope	10 EU countries: Belgium, Germany, France, Italy, Netherlands, Poland, Romania, Spain, Sweden, United Kingdom
Methodology	Online (quantitative survey)
Sample size	N=8,000 (n=800 per country)
Quotas	Country; Gender; Age group;
Sampling error	±1.12% for overall data and ±3.54% for country-specific data. In all cases, a maximum indeterminate probability (p=q=50), for a confidence level of 95.5% is applicable for each one of the reference populations
Sampling	Random

2.2 Descriptive findings

The results of the analysis of the survey data are contained in about 150 tables and graphs, making up the 122 pages of **Annex III,** that the interested readers can consult, whereas in this paragraph we only report the main results in textual forms or in evidences boxes, cross-referencing some of the figures and/or tables contained in the cited annex. In some cases, either in the evidence box title or in the text, we also indicate the exact item (i.e. Q8, Q29, etc.) in the questionnaire (**Annex V**) from which the data are generated.

The large majority of respondents in our sample have either purchased a car or participated in a car purchase decision.

Evidence Box 1 Car ownership / purchase decision

"

- 7.6% of respondents report never having purchased a car (Figure 15 of **Annex III**);
- 6.5% report never having participated in a car-purchase decision, and in 56% of cases the decision is shared with other family members or friends (Figure 13 of **Annex III**);
- Among those who report possessing a car (92.4% of the total sample) about 70% affirms that usage is shared with other family members (Figure 16 of **Annex III**)
- In terms of sharing or not sharing decision and usage the differences among countries do not seem to suggest any possible cultural or geographical patterns (Figure 14 of **Annex III**)

"

Evidence Box 2 Engine types actual and future

"

▶ **Gasoline**:	current car 59.9%	future car: 30.5%
▶ **Diesel**:	current car 42.4%	future car: 38.0%
▶ **Alternative fuel**:	current car 3.2%	future car: 10.9%
▶ **Hybrid**	current car 1.8%	future car: 24.0%
▶ **Electric**	current car 0.1%	future car: 8.9%
▶ **Don't know**	current car 1.2%	future car: 11.5%

"

In terms of the factual and prospective information (Evidence Box 2) it is noteworthy that: a) whereas currently only about 1.9% of the sample possess either a hybrid (1.8%) or an electric (0.1%) car (Q10[11]); b) almost 33% of respondents affirm that in the future they plan to buy either a hybrid or electric car (Q12[12]).

Evidence Box 3 Attributes in order of importance (Q16)

"

Very important:

1)	Price:	59.2%
2)	Fuel consumption:	45.8%;
3)	Safety:	41,6%
4)	Running costs:	35.6%
5)	Size:	28.7%
6)	Engine type:	24.5%
7)	Tax incentives:	20.1%
8)	Lifestyle:	18.3%
9)	Performance:	16.8%
10)	Environmental impact	16.7%
11)	Customisation:	11.7%

"

[11] Total does not add to 100%, as respondents who owned more than one car could choose more than one engine type car when indicating future purchase intentions.

[12] Total does not add to 100%, as respondents were allowed to indicate at least two engine type cars.

As can be seen from Evidence Box 3, among the attributes, taken into account when considering a car purchase, the environmental impact of a car is considered as a very important factor by only 16.7% of the sample, whereas price is considered very important by almost 60% of respondents. On the other hand, the answers to Q16 seem to suggest that factors pertaining to the broadly defined "Fuel Economy" score fairly high in importance: fuel consumption and running costs are considered very important by, respectively, 45.8% and 35.6% of our sample. These findings, however, should be interpreted in combination to those of Q17 (see next evidence box) where we find low percentages of full agreement with statements affirming that environmental parameters are important, and quite high percentage of full agreement with statements affirming that attributes such as price, type of engine, and class of car are already decided before selecting a car model. If this is the case and if consumers decide on type of engine and class of car before selecting a car model then it seems unlikely that they will fully take into account eco-friendliness parameters.

Evidence Box 4 Full agreement with statements from Q17

"

- Full agreement with the following statements is low:
  - I am ready to pay more to guarantee environmental protection for a more ecological model: 8.2% of all sample;
  - Environmental effects (car emissions) are the main factor in defining the current class of car I choose: 9.1% of all sample;
  - Environmental effects (car emissions etc.) may push me to change the size of the car (e.g. from SUV to mid-size car): 11.4% of all sample;
- Full agreement with the following three statements is quite high:
  - Before choosing my car, I already decided the price range: 47%;
  - Before choosing my car, I already decided its price and maintenance costs: 30.6%;
  - Before choosing the car, I already decided its engine type: 25.1%;
  - I select the car among those belonging to a certain vehicle class: 23.8%;
  - Before choosing my car, I already decided its size: 22.2%.

"

In general we can conclude that answers to Q16 and Q17 fully confirm the key points we had encountered in the review of the literature:

- **Two stage purchase process**. There is no doubt that the selection of the class of car is only marginally impacted by eco-friendly and fuel economy parameters, factors that may enter only when a model is chosen within a class;

- **Greater importance to classical attributes**. Price and other classical parameters remain the key attributes considered important by consumers;
- **Inconsistency of self-reported attitudes and intentions**. Answers to Q16 and Q17 seem in clear contrast with the reported intentions (evidence box 1) to buy in the future cars with alternative engine types and with the awareness about the environmental impact of cars that we present later in this paragraph;
- **Potential importance of fuel economy nudges**. It seems that fuel economy parameters are considered as fairly more important than information about CO2 emissions and other environement related items, which may suggest that nudges should focus on these aspects.

Four questions (Q18 through Q21, see **Annex V**) addressed the information sources in general and were specifically aimed at verifying whether the online channel of information was more important than others, in view of possible implications for future policy decisions on *ad hoc* online labels. We can conclude from the data that this channel is important but alongside many other sources of information. More precisely among sources of information for car purchase; Internet, family and friends, and experts score about the same in terms of importance according to the answers provided by our respondents.

In terms of contextual factors it is worth mentioning that only 5.7% of respondents totally agree with the statement "*most people are willing to pay higher prices to protect the environment*". This shows that the faith in the eco-friendly behaviour of others is fairly low, which may reduce the perceived effectiveness of one's behaviour as consumer and induce 'keeping up with the Joneses' attitudes (if other do not care and free ride on the public good 'environment' then why should I behave differently?). The data shows, however, that lack of faith in others does not prevent 61.3% of respondents agreeing with the statement "*my lifestyle can have an impact on the environment*" and 42% to disagree with the statement "*it s hard for someone like me to do much about the environment*". So, despite lack of faith in others, respondents do not feel that this exempts them from their individual responsibility. On the other hand, there is quite a widespread perception that less polluting or lower consumption vehicles are associated with higher prices and that to some extent they also compromise performance. So, we also find evidence of the fact that many consumers perceive buying eco-friendly car as entailing loss/sacrifice in terms of other parameters.

Looking at the three questions (Q29, Q30, Q31, see **Annex V**) which ask about awareness of the environmental impact of car usage, we have a final confirmation of the 'attitudes-action' gap. The sample is almost evenly split between those who consider the pollution attributable to car usage to be below 40% and those who consider it to be above 40%.

About 56% of respondents agree that cars contribute significantly to CO2 emission and its consequences, but 46% also agree that car pollution is not the main source of deterioration of the environment. These results are very robust across the population; in fact, very few socio-demographic and country differences are statistically significant with respect the answer to these questions

Whereas general awareness about the environmental impact of car usage seems fairly high, when asked a question on preferences clearly related to their actual behaviour, the majority of respondents revealed they would rather opt for their own private car for daily commuting than for alternative means of transportation (trains, car pooling, bus, etc.). This confirms the existence of a clear "attitudes-action" gap.

In the final part of the survey respondents were shown examples of existing labels and were presented with several questions to reveal: a) their awareness of the labels; b) their trust in the label; c) the perceived credibility of the labels.

Existing car labels do not score particularly highly since as much as 49% of the sample disagree with the statement that they are familiar with car labels (Q32), almost 40% with the statement that they are easily recognisable, whereas 44.5% agree that car labels are unfamiliar to them. Many also mistake car eco-labels as an indication of product reliability. On the other hand, perceived credibility is fairly high, although many think that the information contained in the label is not sufficient. When asked to react to statements concerning how they use labels when they buy a product (any product and not a car), very few agree that they never look at the information on the label (14.2%). On the other hand, it appears that careful inspection of the labels is conducted only when buying a product for the first time.

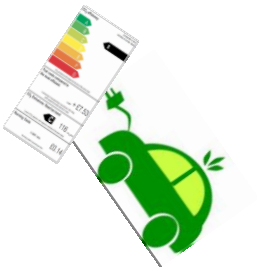
2.3 Split ballot findings

As illustrated earlier, at the end of the survey (after item Q36; **Annex V**) the sample was split into four ballots for a preliminary and only exploratory test of the relative effectiveness of four CO2 classification systems (see graphic illustration in Figure 17, p. 99). This only purpose of this test was to inform the subsequent design choices of the laboratory experiment and should not be taken as conclusive. The four measures used for the test are self-reported scales of the noticeability effect the labels had on the respondents' perception. It does not matter whether the answers to the four questions correspond or not to the objective parameter of the car respondents had seen. What matters is whether we can identify systematic differences in the way the four

classification systems shape the answers to the relevant questions (items Q37a,Q37b, Q37c, Q37d; **Annex V**).

To this purpose we performed ANOVA (Analysis of Variance) between the split ballots. In other words, we tested whether there was any statistically significant difference between the ballots (four sub-samples each associated with one of the four classification systems) in the way respondents answered the mentioned four questions.

It is important to add that the absolute classification system is used as the baseline against which the others are compared. The findings unequivocally show that the relative classification system induces systematically higher evaluation in terms of environmentally friendliness, fuel efficiency, with also a smaller variance as compared to all other three systems. At the same time, the lowest evaluations were elicited by a specific sub-variant of relative rating- the German classification system (with the highest variance), while the 'combined' one did not have any particular effect compared to the benchmark. The preliminary conclusion would be that the relative classification system is more effective in producing a noticeability effect. However, this effect might have been provoked by the efficiency class that each rating indicated for a variant of the VW Polo used in the survey (i.e. according to the German system the car was rated as class D, the relative system resulted in class B, absolute and combined in C) As stated, this was a preliminary exploration and the effect of the classification systems were tested again in a more comprehensive way both in the laboratory and in the online experiments.



3 Laboratory experiment

3.1 Design

The full illustration of the design, treatments, randomisation procedures and protocols, sample, and response variables used can be found in the Technical Compendium (chapter four). In the Compendium the reader will also find synoptic tables of all treatments and of all response variables (at the beginning of the document).

Additionally the pre and post treatment questionnaire and the experimental tasks[13] are reported in **Annex VI**. Finally in the Appendix of this report all the treatments tested are listed and exemplified graphically (§ 6.2).

Hence, below we very succinctly recall the main aspects of the laboratory experiment design.

It was a randomised control trial based on between-subject design with repeated measures whereby respondents were allocated either to a treatment (nudge) or the placebo (control condition) but for statistical reasons were exposed to more than one experimental condition[14].

A very sophisticated randomisation was designed whereby not only subjects were randomly allocated to experimental conditions, but the order by which they were exposed to these conditions was also randomised, and the cars they were asked to buy were also extracted randomly from our car database (see more on this chapter 6 of the Technical Compendium).

[13] The pre and post treatment questionnaire contained in Annex VI are exhaustive, but for the experimental tasks we only report three examples among the many that could results from the randomisation loops.

[14] In pure between-subject design subjects are either allocated to one treatment or to the control condition.

The control condition was the car label currently in use in the UK (Figure 18, p. 99) The choice of the UK label was simply driven by the fact that the Laboratory sessions were run in London. The tested treatments (nudges) were the following:

Common standard elements:

1. **Graphic layout of the CO2 classification systems, two values**: horizontal; vertical (Figure 19, p. 99);
2. **CO2 classification systems, three values**: absolute; relative; combined (Figure 17, p. 98);

Common additional elements

3. **Running costs, three values**: per mile; monthly; per five years (Figure 20, p.100);
4. **Lost saving on fuel, present or absent**: formulated in terms of lost saving per mile compared to best (most fuel efficient) vehicle in class (Figure 21, p. 101);

Additional element specific for conventional engine cars:

5. **CO2 related taxation, present or absent**: formulated in terms of level of taxation compared to best (with lowest CO2 emissions) vehicle in class (Figure 21, p. 101);

Additional elements specific for electric engine cars:

6. **Tailpipe and upstream emissions, two values**: tailpipe only; tailpipe and upstream (see Figure 22, p. 102);
7. **Consumption, two values**: Range in distance covered with fully charged battery; electricity consumption (see Figure 22, p. 102);

Additional elements specific for hybrid engine cars:

8. **Tailpipe and upstream emissions, two values**: tailpipe only; tailpipe and upstream (Figure 23, p. 102);
9. **Consumption, two values**: separate figures for fuel and other sources; one synthetic indicator for all sources (Figure 23, p. 102);

Before moving forward we must further delve into the issue anticipated at the end of § 1.1 regarding the practical definition of a treatment.

The reader can appreciate the very large number of conditions that were requested to test. The conditions should not be summed up, but multiplied, because of all the possible interactions. A simple calculation would show that if we wanted to test the full set of combinations of all the elements above, we would have to allocate participants to 72 different combinations of treatments plus control. These would be 72 different labels tested as the treatments. Since we know that, as rule of thumb, we need to have at least

60 participants (30 for treatment and 30 for control) for each randomisation run (see § 4.3 of Technical Compendium), this means that we would have needed a sample defined as (72*30) = 2160 participants, which is beyond feasibility in any laboratory experiment. This is why we had to implement a main effect design and not a full factorial design. In practice this means we are not testing the labels as such (as a specific combination of various pieces of information with their formats) but only the various information elements they contain.

Let us try to be even more concrete taking one example and explaining the data and how it is processed, in the regression analysis, by the statistical software package. Let us consider the condition vertical graphic layout and the response variable noticeability effect of environmental friendliness (measured by answers to a scale 1 to 10).

The independent variable is a dummy equal to 1 if respondents saw the treatment vertical layout and equal to zero if otherwise. The dependent variable are the scores 1 to 10 expressed by the respondents on the eco-friendliness of the car they selected to buy. By means of a statistical test ('regression analysis'), we compare the outcomes in terms of reported eco-friendliness of the car purchased respectively between those who see the label, with the vertical layout, and those who did not. If this difference is positive and large enough to avoid being explained by pure chance (statistical significance) then it can be taken as an empirical estimation of the causal effect. Statistical theory tells us that for a sufficiently large sample, this estimation is equal to the true value of the causal impact.

The regression analysis is blind, both with respect to the visual stimuli as a whole and with respect to other elements contained in it other than the graphic layout treatment. It processes only the single nudge and all the scores for the dependent variable. This is what having a 'main effect' design rather than a 'full factorial' design means.

We should stress that we are not stating that interactions do not matter, but that we had a constraint in terms of feasibility. It was agreed since the very beginning (kick off meeting) with DG CLIMA that the policy requirement to assess all these nudges should be traded off with the lack of analysis on the possible interaction effects.

Finally, for the laboratory experiment we used a sample recruited from the panel managed by London School of Economics and Political Science for its behavioural lab. As common in most of the experimental literature, this sample is not representative of the general population.

3.2 Main findings

The technical tables with all the detailed results from the regression analysis are reported in § 4.5 of the Technical Compendium, whereas below we report the main findings plainly and mostly in textual form (with the addition of a few boxes).

3.2.1 Noticeability effects

Conventional cars. We had several different treatments whose effect is measured in terms of corresponding response variables. For instance we tested whether the information on lost savings on fuel led participants to score higher (compared to those who were exposed to the control condition) the car they selected when answering the question "how do you think the car you selected score in terms of running costs compared to the other options available?" (scale 1 to 10 where 1 means 'much worse' and 10 'much better'). So, we regressed: a) the eco-friendliness noticeability response variable (also abbreviated as "**NoticE**") against both graphic layout (two levels: vertical and horizontal) and classification system (three levels: absolute, relative, combined) treatments; b) the fuel-efficiency noticeability response variable (also abbreviated as "**NoticFE**") against the 'Information on Lost Saving on Fuel' treatment (single level); c) the CO2 taxation noticeability response variable (also abbreviated as "**NoticT**") against the homonymous treatment (single level); and d) the running costs noticeability response variable (also abbreviated as "**NoticRC**") against the homonymous treatment (three levels). The key results are reported in the box below.

Evidence Box 5 Noticeability effects (conventional cars)

"

- Both vertical and horizontal layout have statistically significant positive noticeability effect on **eco-friendliness**;
- Relative and Absolute classification systems have statistically significant positive effect on **eco-friendliness**, with the former seemingly stronger, whereas the combined system has no effect;
- There is no noticeability effect of the information on lost saving on fuel on **fuel efficiency** and of CO2 taxation on **CO2 related taxation**
- Only **running costs per month** have a positive and statistically significant noticeability effect.

"

The positive and statistically significant effect of both the vertical and graphic layout means that those respondents who saw either of them, on average, rated the car they selected to purchase as more eco-friendly than the respondents who where exposed to the control condition rated the cars they selected to purchase. Indeed, one may wonder how this is possible, since the control condition also contained a graphic layout. The answer is embedded in the explanation we provided at the end of the previous paragraph about the difference between a main effect and a full factorial design. In other words, there must be some added value of the new nudges that were shown together with the graphical layout, but we cannot exactly identify them due to statistical constraints embedded in the design.

On the other hand, the statistical test (pairwise comparison among significant coefficient) cannot establish which of the two (vertical or horizontal) is more effective, i.e. we cannot definitely state that the difference among the two is due to causal difference (one being more effective) or simple chance. With regard to the classification system, the combined one turns out not to be significant and the relative one is arguably the more effective, which seems in line with the results of the split ballot. For all other 'additional information' treatments (information on CO2 taxation, lost saving on fuel, and running costs), and their corresponding noticeability measures, we mostly find no significant effect, except for the information on running costs framed as "running costs per month".

Electric and hybrid cars.We proceeded in the same way as above but we do not report the various variables since we find no statistically significant and noteworthy noticeability effect of any treatment on any of these variables.

3.2.2 Cognitive effects

Conventional cars. As anticipated, for "comprehension and recall" of information, we triangulate the evaluation provided by respondents on the car they selected to purchase (answering the post treatment questions) with the objective information contained in our database and determine (using a range, see § 4.2 of Technical Compendium) whether the evaluations are correct or incorrect.

Since these variables are not immediately easy to understand, we provide an example. Let's take the case of a respondent who has performed a simulated purchase from three cars of a certain class and then reported that the chosen car scored ten in term of running costs compared to the other options available (i.e. the car provides a significant amount of savings due to running costs). Implicitly, she is stating something: "I judge the car to be the one that has better performance in terms of running costs". If the number is below 6.67 (two thirds of ten) and above 3.33 (one third of ten), we can reasonably assume that

she judges the car to be the second best scorer; if the score is below 3.33 then she is scoring it as the least performer. Since we have the three cars recorded in the log file of the experiment and since we have the information from the database, we could easily compute the correct ordering of cars and compare it with the 'revealed' one explained above. Our variable is a dummy equal to one if the two match, zero otherwise. This means 1 if the judgement by the respondents is confirmed by objective information as correct. Since we take into account cognitive limits and computations errors, we also allow them to be mistaken by one position. In other words we care only if a revealed "top performer" is at least a second best.

We, thus, constructed four dummy variables (1 if almost correct recall and 0 otherwise)that are regressed against the treatments to test whether respondents retain and correctly recall the information conveyed or not. The four dummy variables with their abbreviations are: a) cognitive measure of recall of the CO2 emission information ("**CEM"**); b) Cognitive measure of recall of the fuel efficiency information ("**CFE"**); c) Cognitive measure of recall of the CO2 emission related tax information ("**CCO2T"**); and d) Cognitive measure of recall of the Running Costs information ("**CRC"**) From the regression analysis we find no statistically significant effect, except for the treatments "running cost per month" and "running costs per five years", but we cannot conclude which of the two is more effective in terms correct information recall. This means that in most cases when respondents were asked to score the car they selected in terms of emissions, costs, etc. they answered in ways not significantly different from the control conditions. In other words they do not seem to recall the information in the label in a more or less effective way compared to the respondents who were exposed to the control condition (represented by the standard car label in use in the UK) This may also mean that, despite the fact that our visual stimuli of the labels were less overloaded and clearer than the UK label, they nonetheless contain too much information for consumers to retain and correctly recall.

Electric cars. We proceeded in the same way as explained for conventional engine cars and constructed the following dummy variables with their abbreviations: a) Cognitive measure of recall of the Fuel Efficiency information ("**CFE"**); b) Cognitive measure of recall of the Running Costs information ("**CRC"**); c) Cognitive measure of recall of the Electric Consumption information ("**CEC"**).We find positive and significant effects for all three formats by which running cost information is presented across all the response variables, and for electricity consumption information with respect to the recall of fuel efficiency information.

Hybrid cars. We proceeded in the same way as explained for conventional engine cars and we constructed the following dummy variables: a) Cognitive measure of recall of the Fuel Efficiency information ("**CFE"**); b) Cognitive measure of recall of the Running Costs information ("**CRC"**); c) Cognitive measure of recall of the Combined Consumption information ("**CCCI"**). The only effective nudge is 'lost saving on fuel', which helps respondents in correctly rating the car they selected in terms of fuel efficiency.

3.2.3 Behavioural effects

Conventional cars. We used the simulated purchase choice and the self-reported budget respondents intended to spend on their next car (declared in the pre-treatment questions, so before the task starts) and we checked whether, as a result of the nudges, their willingness to pay changed. In the sense that they may buy a car that costs more than their declared budget as this is reflected in future savings on fuel or lower CO2 emissions. So, we actually use as a measure the variable Change in Willingness to Pay (**CWTP**). We find, however, no significant effect for the treatments tested, with the only exception being the horizontal graphical layout where we see a positive effect, statistically significant at 10% level.

We then constructed two additional variables using the purchase choice as an indicator of green behaviour. The variables are scores (1 to 10) for the choices that respondents made compared to the other options of car they were shown but did not select to buy.[15]

One is a score concerning the emissions, whereby the higher the score the lower the emissions of the car chosen by respondents, and so the greener the choice. For this variable we also use the abbreviation "**B Emis"**.

The second score, calculated with the same logic as the previous one, is about the lost savings on fuel consumption, where the higher the score, the higher the savings, and the greener the choice. For this variable we also use the abbreviation "**B LSF".** Main results are reported in next box.

The fact that we find no effect for emissions but some for lost savings on fuel seems to suggest that information about tangible monetary measures has a more direct behavioural effect than information about CO2 emissions.

[15] More precisely, they are the scores used to compare with the revealed ones in the construction of cognitive variables described above.

Evidence Box 6 Behavioural choices effects (conventional cars)

"

- For "**B Emis**" we do not detect any statistical significant effects for the treatments used compared to the control condition;
- On the contrary, for "**B LSF**" we find several positive and statistically significant effects for the following treatments:
  - Both graphic layout (horizontal and vertical);
  - Combined classification system;
  - Information on lost saving on fuel, as compared to the most efficient car in its class;
  - Running cost per mile;
  - Running cost per five years

"

Electric cars. The response variable Change in Willingness to Pay was constructed using the same logic as per the conventional engine cars. On this response variable we find statistically significant positive effects only for the treatments 'Battery Range' and 'Running Cost per 5 years'[16].

As done with the conventional engine car we construct the variable "**B LSF"** and find quite interesting results with several statistically significant positive effects. The two graphical layouts have a positive and significant effect, but we cannot distinguish statistically among them. Both the indication of tailpipe and upstream emissions has a positive effect, but the second is stronger and statistically different from the first one. Finally, all the three running costs formats have a positive and significant impact, but the three are not statistically different from each other.

Hybrid cars. The response variable Change in Willingness to Pay was constructed using the same logic as per the standard engine cars. On this response variable we find no appreciable statistically significant effect.

As done with the conventional engine car we construct the variable "**B LSF"**and the results are quite interesting with several statistically significant positive effects. The Graphical 1 layout has a positive and significant effect in nudging towards more "fuel efficient" cars. The same happens for both tailpipe and combined tailpipe and upstream

[16] However, we should keep in mind that the budget was declared for buying a conventional engine car, while everybody has to perform an additional task for alternative engine. This caveat is important in the interpretation of these results for alternative engine cars. Further details are provided in the Technical Compendium.

emissions, but the latter has a stronger and statistically different impact. Combined (electricity and other fuel) and separate consumptions figures are both effective, but the latter has a stronger effect. Finally, for the running costs the two best performers are the monthly and the five years ones, but they are not distinguishable statistically.

3.3 Synoptic summary

We summarise in the three tables below the main findings of the analysis reported in the previous sub-paragraphs, where we highlight (light green) the cells with positive and statistically significant effects. The rows report the treatments (grey rows indicate treatments with no effect on any response variables) and the column the different response variables. We remind that all results are produced testing the treatments against the control condition.

Table 3 Synoptic overview of laboratory experiment results (standard)

	Notic	CWTP	B Emis	B LSF	CEM	CFE	CCO2	CRC
Graphic vertical	Y	N	N	Y	N	/	Y	/
Graphic horizontal	Y	Y	N	Y	N	/	N	N
Absolute	Y	N	N	N	N	/	/	/
Relative	Y	N	N	N	N	/	/	/
Combined	N	N	N	Y	N	/	/	/
Lost savings on Fuel	N	N	N	Y	/	N	/	/
CO2 Tax	N	N	N	N	/	/	N	/
Running costs: mile	N	N	N	Y	/	/	/	N
Running costs: month	Y	N	N	N	/	/	/	Y
Running costs: 5 years	N	N	N	Y	/	/	/	Y

Note for all three tables: Y= positive and statistically significant; N= no significant effect or negative effect;/= cells where there is no regression result for response variables (column) were not relevant for the treatments (rows) were not regressed. CWTP= Change in Willingness to Pay. Notic= noticeability and it is assumed that it refers to different response variables in relation to the different treatments in the rows.

Table 4 Synoptic overview of laboratory experiment results: (electric)

	Notic	CWTP	B LSF	CRC	CFE	CEC
Graphic vertical	Y	N	Y	/	/	/
Graphic horizontal	N	N	Y	/	/	/
Tailpipe	N	N	Y	/	/	/
Tailpipe & upstream	N	N	N	/	/	/
Lost savings on fuel	N	N	N	/	N	/
Electric consumption	N	N	Y	/	Y	N
Battery range	N	Y	N	/	N	N
Running costs: mile	N	N	Y	Y	Y	/
Running costs: month	N	N	Y	N	N	/
Running costs: 5 years	N	Y	Y	N	Y	/

Table 5 Synoptic overview of laboratory experiment results (hybrid)

	Notic	CWTP	B LSF	CRC	CFE	CCCI
Graphic vertical	Y	N	Y	/	/	/
Graphic horizontal	Y	N	N	/	/	/
Tailpipe	N	N	N	/	/	/
Tailpipe & upstream	N	N	Y	/	/	/
Consumption Info: separate	N	N	Y	/	N	N
Consumption info: combined	N	N	Y	/	N	N
Lost savings on fuel	N	N	N	/	Y	/
Running costs: mile	N	N	N	N	/	N
Running costs: month	N	N	Y	N	/	N
Running costs: 5 years	N	N	Y	N	/	N

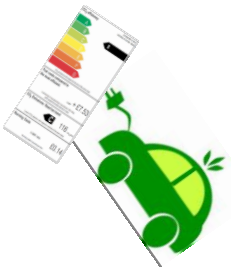
As can be seen from coloured cells in the three tables, we have found several significant effects and, thus, the first screening has not been inconclusive. Moreover, we could already "eliminate" some of the information that has proved to be ineffective. On the other hand, it is clear that the results are not consistent across treatments, measures, and engine types, but somewhat scattered across them. In other words, we cannot draw clear-cut conclusions on the most effective treatments, especially since pairwise comparisons of effective treatments do not yield conclusive results, except in the case of noticeability and classification systems where the Relative solution appears more effective than both Absolute and Combined.

On the other hand we find a number of significant effects for nudges falling within the broad label of 'fuel economy':

- For conventional engine cars:
  - Information on fuel lost savings
  - Running costs per mile
  - Running costs per 5 years
- For electric cars:
  - Information on electric consumption
  - Running costs per mile
  - Running costs per month
  - Running costs per 5 years
- For hybrid cars:
  - Combined information on consumption
  - Separate information on consumption
  - Running costs per month
  - Running costs per 5 years

In addition battery range and running costs per 5 years, in the case of electric, have a significant effect also on Change in Willingness To Pay (CWTP). We can, thus, conclude that fuel economy parameters work better than more strictly defined environmental information (CO2 emissions and related taxation)

We could also draw two implications that to some extent subsequently shaped the design of the online experiment. First, with respect to the three classification systems for standard engine cars we do not find enough evidence to conclude which is the most effective among absolute, relative, and combined. Therefore, we tested again the classification systems through a specific sub-task of the online experiment. Second, and most importantly, the labels tested in the laboratory seemed to suggest a risk of overload, which asks for a re-test of the labels in the online experiment but using simplified versions.



4 Online experiment

4.1 Design

The full illustration of the design, treatments, randomisation procedures and protocols, sample, and response variables used can be found in the Technical Compendium (chapter five). In the Compendium the reader will also find synoptic tables of all treatments (at the beginning of the document, of all response variables (Table 44), and of all the kind of analysis performed (Table 46). Additionally the pre and post treatment questionnaire and the experimental tasks[17] are reported in **Annex VII**. Finally in the Appendix of this report all the treatments tested are listed and exemplified graphically (§ 6.3). Hence, below we very succinctly recall the main aspects of the laboratory experiment design. It was a randomised control trials based on between-subject design whereby respondents were allocated either to a treatment (nudge) or the placebo (control condition). We re-tested the elements in the labels but in simplified versions (compared to the laboratory experiment) and we tested the promotional material. We exploited the large sample size and split it into sub-samples as to both re-test labels and test promotional material without placing too much burden on respondents and, especially, staying within the 25 minutes duration that is considered in the literature the maximum acceptable duration of an online experiment before noise kicks in (respondents are tired and provide unreliable answers). Hence, respondents were randomly allocated to:

- Label sub-task 1.1 (N=800): classification system;
- Label sub-task 1.2 (N=3200): further testing of labels;

[17] The pre and post treatment questionnaire contained in Annex VII are exhaustive, but for the experimental tasks we only report three examples among the many that could results from the randomisation loops.

- Promotional material sub-task 2.1 (N=1200): test with chosen format (same treatments as other sub-tasks, but they can select in which format the promotional material was shown to them);
- Promotional material sub-task 2.2 (N=2400): test with random format;
- Promotional material sub-task 2.3 (N=400): test with process tracing (same treatments but usage of blurred image as illustrated and explained earlier in § 1.3, see Figure 7, p. 34)

The condition tested were the following:

Task 1 re-test of labels:

1. **CO2classification systems, three values:** absolute, German, full label, tested only with conventional engine cars
2. **Control condition:** standard label with vertical CO2 absolute classification
3. **Running costs, two values:** cost per mile/km, cost per 5 years (for conventional and hybrid cars see Figure 26; for electric see Figure 27)
4. **Lost savings on fuel[18], two values:** framed as additional costs, frame with text "you lose" followed by the additional costs (see Figure 26);
5. **Fuel economy, present or absent:** litres per km or miles per gallon depending on the country, for conventional engine cars and for hybrid cars(see Figure 26);
6. **Savings on fuel for electric car:** expressed with the text "you save" and the money value, savings calculated comparing with the most efficient conventional engine car in the same size class (Figure 27), alternatively as the treatment four above, i.e. additional costs with respect to the best 'electric' car in the database;
7. **Battery life for electric cars, present or absent.**

Task 2 promotional materials:

1. **Control condition**: format in which they are currently available on the market (Figure 28);
2. **General format, two values:** information about CO2 emissions in graphic format, information about CO2 emissions both in graphic format and in text format(see Figure 29):
3. **Additional Element, three values:**containing only the CO2 emission class (also simply Class), containing the CO2 emission class plus a small text indicating running costs (also simply RC small), containing the CO2 emission class plus a larger running cost element (also simply RC salience; see Figure 30);
4. **Web link[19]:** Present or absent

[18] As compared to the most efficient model in a class.

The online experiment, exactly as the preliminary survey, is based on a sample including 800 individuals per country in the 10 countries selected for a total of 8000 respondents. The countries included are: BE=Belgium; DE= Germany; FR= France; IT= Italy; NL= Netherlands; PL= Poland; RO= Romania; ES= Spain; SE= Sweden; UK= United Kingdom. The main technical parameters of the sample are reported in next table.

Table 6 Sample (online experiment): summary technical parameters

Population	General population aged 18 to 65 years old
Scope	10 EU countries: Belgium, Germany, France, Italy, Netherlands, Poland, Romania, Spain, Sweden, United Kingdom
Methodology	Online (quantitative survey)
Sample size	N=8,000 (n=800 per country)
Quotas	Country; Gender; Age group; Experience in purchasing cars
Sampling error	±1.12% for overall data and ±3.54% for country-specific data. In all cases, a maximum indeterminate probability (p=q=50), for a confidence level of 95.5% is applicable for each one of the reference populations
Sampling	Random

[19] The web link leads to a website where a full label for the car is visualised.

4.2 Main findings: re-test of labels

4.2.1 Ranking task: absolute versus German

In this case respondents saw three cars, vertically disposed with a label (the order of presentation was randomized) and they were asked to rank them in terms of growing emissions. The cars were randomly selected from a subset of our database (three or four cars per class). With 50% probability the label included only the classification system (and it was equally likely to include a German or an Absolute classification), while with 50% probability the label contains a full set of information. The comparison of the results of the former (only classification system) versus the latter (full label) is a test of information overload. Using our car database we could check how close to the real value was the ranking performed by the respondents. The results are summarised in the box below.

Evidence Box 7 Absolute versus German

"

- The German label is clearly confusing with respect to the absolute classification since on average the score with this system is less correct than the score with the absolute classification (coefficents are statistically significant and negative);
- The absolute classification is immediately clear and effective (more correct answers);
- Even if the absolute classification is presented in combination with other information (as in a full label), there is no cognitive pressure. The hypothesis of information overload (for the case of CO2 emission information provided through absolute classification) is rejected by the data;

The results reported above are confirmed by all technical checks performed and hold when including stratification variables (gender, age, country).

4.2.2 Willingness to pay

With the Multiple Price List respondents were shown a car in combination with the visual stimuli containing the treatments and were asked to express the price they would be willing to pay. The list of prices (the grid) was centred on the current market price of that specific car in the specific country were the subject is answering the questionnaire. Each individual point of the grid has the same increase in each country (6% variation with respect to adjacent prices, in increasing order of magnitude) except for rounding up procedure due to the monetary scales. When we compute the responses we record the variations in prices (e.g. +6%, -12%) with respect to the centre of the grid, controlling for the specific cars and the differences across countries.

With this variable we assess whether the various nudges tested increase or reduce the amount of money individuals are willing to pay. Please remember that the reduction or increase is considered with respect to the control condition or counterfactual (the standard label with no additional information). The interpretation of the decrease or increase in WTP naturally depends on the kind of treatments one is considering. For instance, if information on CO2 emission reduces WTP this may be interpreted as indicating that this nudge induces individuals to reflect on the environmental consequences of that specific car and, thus, express a lower willingness to pay. Overall the treatment tested do not seem to have any significant effect on WTP except for an isolated effect of running cost per mile/km in the case of electric car (see box below).

Evidence Box 8 Willingness to Pay for Labels

"

- We find no significant results, with the exception of running costs per km/mile in the case of electric cars;
- In this case the inclusion of running costs per mile/km decreases the willingness to pay;
- This may be consistent with a positive information role of this format for running costs associated with a valence bias by consumers;
- Valence occurs when the agent tends to overestimate good events, which in this case may be related with real savings from usage of electric vehicles

"

The result of running costs per mile reducing the willingness to pay for electric car is a counter-intuitive result that we will discuss together with others in the conclusive chapter 5.

4.2.3 Noticeability variables

The logic of the noticeability variable is the same as that explained for the Laboratory experiment, but for the ease of the readers we briefly recall it here. Noticeability is measured through a series of direct questions that explicitly mention the information appearing on the visual stimuli. For example, since respondents are exposed to the running costs in two different formats, we then asked them to evaluate the performance of the car they selected in terms of running costs and we can test how much they "noticed" the information. More concretely, we ask them "How do you think the car you selected scores in terms of running costs compared to the other options available?" and we can appraise how much they noticed this parameter by the answers they provide. The noteworthy results are reported in the box below and are confirmed in all technical checks performed and hold when controlling for gender, age groups and country.

Evidence Box 9 Noticeability variables for Labels

"

- Conventional engine cars:
  - Running costs per mile/km increase noticeability with respect to running cost per 5 years, in terms of fuel efficiency, fuel consumption and running costs performance;
  - Lost saving on fuel is more noticeable when formulated as "you lose X with respect to the best performance car" (the effect is statistical significant in terms of noticeability of fuel consumption performance);
  - Fuel efficiency is effective in increasing noticeability (the effect is statistical significant in terms of noticeability of fuel consumption performance);
- Electric cars:
  - Running costs in the format per mile/km is more effective in terms of noticeability, for both electricity consumption and running costs

"

4.2.4 Cognitive variables

Also for cognitive variables the logic by which these scales have been constructed is the same as we have illustrated already for the laboratory experiment, but it is worth to briefly recall it here. We compared the answers of respondents scoring the cars they have

seen on most of the information contained in the visual stimuli with the exact value contained in our car database. In other words, when they provide an answer to the question "How do you think the car you selected scores in terms of fuel consumption with respect to other cars in the market?" they 'reveal' a certain ordering (let's call it ordering 1). We can estimate the real ordering through the database (ordering 2). Ordering 1 and 2 are reported in the same scale, from one to ten. If the difference among the two is less than 25% (leaving the subjects a margin of error), then our cognitive variable is equal to one, zero otherwise.

A correct answer can be interpreted as indicating that respondents are able to retain and correctly recall the information to which they have been exposed, which would be a positive result from a policy perspective. Otherwise, if the answers are far away from the true value then this means that respondents have difficulty in retaining and correctly recalling the information. A few incorrect answers could not justify the conclusion that the information nudges are difficult to retain and recall. On the other hand, if incorrect answers are spread across most or many of the nudges and response variables then, as in the case of the laboratory results, one may reasonably infer that there is a problem of information overload. Since in the online experiment we simplified the contents of the visual stimuli we would expect to find more significant results for cognitive variables than was the case for the lab, which is not the case (see box below)

Evidence Box 10 Cognitive variables for Labels

"

- Conventional engine cars:
  - Here the running costs per mile/km do not seem to improve cognitive processing for fuel consumption performance. The result is confirmed by all technical checks and hold when controlling for age, gender, country;
- Electric cars:
  - Lost Saving on Fuel - presented as additional costs with respect to best electric car- increase the cognitive performance in terms of ranking of electricity consumption. The result is robust across specifications;
  - Battery increases the cognitive performance in terms of ranking of electricity consumption.

"

It must be recalled that the results are always obtained comparing treated and control, so the first bullet point, in the box above means that the control condition is more effective than our treatment in helping respondents correctly rate the car in terms fuel consumption performance. While this is just a single and scattered result on the basis of which we cannot draw a sweeping generalisation, the implication would be that less information is more effective than more (running costs items are not contained in the control condition visual stimulus).

4.2.5 Results with process tracing

As explained when commenting the example of how a visual stimulus would appear using process tracing (Figure 7, page 34), this technique enables us to automatically capture information that can be considered as a behavioural measure of what kind of information elements respondents were most interested in. They have been shown a blurred visual stimulus, after being instructed that to fully visualise its elements they needed to move the mouse over them. We, thus, recorded these mouse-over activities and we considered, in the regression analysis, three variables: a) total number of visualisations; b) at least one visualisation; c) full visualisation of all elements. The results are reported in the box below. We see that it seems that only nudges related to broadly-defined 'fuel economy' attracted more interest in statistically significant ways.

Evidence Box 11 Process tracing for Labels

"

- Running costs per 5 years have a positive and significant effects on the number of visualizations of the blurred images;
- Lost saving on fuel presented as additional costs has a positive and significant effect on the probability of complete visualization of the label;
- Fuel economy has a positive and statistically significant effect on the number of visualizations

"

4.3 Main findings: test of promotional material

4.3.1 Willingness to pay

With no further explanation (same considerations developed for the same sub-paragraphs concerning the labels apply here) we report in the box below the main results. We put emphasis on the last statement in the box for we will come back and comment it in the final chapter 5.

Evidence Box 12 Willingness to Pay for promotion material

"

Significant results are obtained only for the sub-sample of respondents who selected the format (and then had to deal with the process tracing technique;

- For conventional vehicles,running costs in small format increase the willingness to pay;
- With selection of the format, the presence of process tracing, in the post-selection task, increases the willingness to pay. A possible interpretation of this effect is the following:
  - Having to deal with mouse-over, in a context in which they had to choose the format in the prior step, respondents made an extra effort to read and cognitively process the information;
  - This more careful examination of the label may then have shaped the evaluation they performed and the choice they made;

"

4.3.2 Noticeability

With no further explanation (same considerations developed to the same sub-paragraph concerning the labels apply here) we report in the box below the main results.

Evidence Box 13 Noticeability variables for promotional material

"

- Promotional material (format randomly allocated):
  - The inclusion of CO2 classification reduces the self-reported perception of environmental friendliness and fuel efficiency;
  - The inclusion of Running costs information (image or note or text) reduces self-reported perception in terms of: running costs themselves, environmental friendliness, fuel efficiency;
  - The simple inclusion of a weblink to a full label reduces self-reported perception of environmental friendliness and fuel efficiency
- Promotional material (format selected by respondents):
  - Information about running costs reduces self-reported noticeability measures in quite a systematic fashion: a) running costs salience reduces all the three measures of noticeability ; b) running costs in small format are not effective;
  - Even the simple footnote on the running cost measurement scale (which accompany the RC information regardless of the format) is less effective than control condition on (self-reported) environmental friendliness and fuel efficiency
  - Text for CO2 is reducing in a statistically significant manner the three self-reported measures.

"

All the results in the table above, some of which may seem counter-intuitive, should be read in view of the explanation provided at the end of § 3.1 about the implications of having a main effect design rather than a full factorial one. This is particularly relevant for the results of the 'web link treatment'. Out of 2789 respondents, exposed to this treatment, 362 (13% of the total) clicked on the web link. But in our regression analysis we did not test whether respondents clicked or not, since this would not have allowed us to recover any causal effect: the web link was not present in the control condition and we could not compare means in the number of clicks between treated and non treated. So,

we used the 'web link' as all other treatments.[20] We can replicate with some adaptation the same hypothetical exemplification presented in § 3.1. Assume that the treated group is made up by those respondents exposed to the promotional material with a weblink and the control group is made up by those who saw the control (no web link). Assume the dependent variable are the scores 1 to 10 expressed by the respondents on the fuel efficiency of the car they saw in the MPL task. We simply looked at the difference in outcome variable (reported fuel efficiency) between the treated and the control group (through a statistical technique called regression)and test if the difference in fuel efficiency noticeability can be due to chance or not. We find a statistically significant coefficient but with a negative sign, meaning the effect of the control condition, in making fuel efficient noticeable, is higher than the effect of the 'web link'.

[20] The web link treatment is the presence of the hyperlink in the promotional material. The hypothesis tested is whether or not its visual presence generates any difference in response by the individuals as compared with promotional material in which this hyperlink is absent. By difference in response we mean that the answer given to specific questions (noticeability), the reasoning done after visualizing the label (cognitive) or the conduct performed after the visualization (behavioural choice). This is independent from the concrete use of the hyperlink, namely the click on it, since being absent in the control condition (by definition) we cannot estimate a counterfactual to compare with the treatment. The hypothesis of absence of effect cannot be rejected in terms of behavioural response. We found a significant difference when we look at noticeability and cognitive variable, but the effect are the opposite: in one case (noticeability, answer to the fuel efficiency question) the effect is negative, while when we look at cognitive processing (score in terms of fuel consumption) it is positive. Moreover, the absence of effect is confirmed throughout the set of measurements when we restrict the sample to people with tertiary education.

4.3.3 Cognitive variables

With no further explanation (same considerations developed to the same sub-paragraph concerning the labels apply here) we report in the box below the main results.

Evidence Box 14 Cognitive variables for promotional material

"

- Promotional material with format randomly allocated:
  - Running costs salience increases cognitive processing for both environmental friendliness and running costs;
  - CO2 classifications image increases cognitive processing for both environmental friendliness and running costs;
  - Weblink increases cognitive performance of information for environmental friendliness
- Promotional material with format selected by respondents:
  - Running costs salience increases cognitive performance of information for environmental friendliness;
  - Running costs in small format decreases cognitive performance of information of running costs;
  - Running costs footnote accompanying the two treatments above increases cognitive performance of information for environmental friendliness;
  - CO2 classifications text increases cognitive performance of information for both environmental friendliness and running costs;

"

All results reported above are confirmed by technical checks and hold when controlling for age, gender, and country.

4.3.4 Process tracing

With no further explanation (same considerations developed to the same sub-paragraph concerning the labels apply here) we report in the box below the main results for the regression of process tracing measures.

Evidence Box 15 Process tracing for promotional material (selected)

"

- Running costs salience has: a) a positive effect on the number of visualizations; b) positive effect on the probability of visualization of at least one part of the promotional material;
- Running costs in small format have a negative and statistically significant effect on both: a) the number of visualizations; and b) the probability of visualizing the full promotional material
- The footnote on Running costs accompanying the two treatments above and indicating the unit of measurement has a positive and significant effect on the number of visualization and a negative and significant effect on the probability of visualizing the full promotional material;
- When we compare positive and significant treatments among them, we cannot discriminate in a statistically significant way, i.e. we cannot say if one is robustly more effective than the others
- CO2 text has: a) a positive effect on the number of visualizations; b) a positive effect on the probability of visualization of at least one element.

"

4.4 Synoptic summary(treatment effects)

In the next three tables (Table 7-Table 9) we provide a synoptic synthesis identifying statistically significant results for the various treatments across the response variables for both labels and promotional material. It is easy to see from the response variables (reported in full and not with the abbreviation) that the rows 1-7 concern labels and rows 8-13 concern promotional material in all three tables. Cells are empty under absence of any noteworthy and significant results. We report the "Not applicable" when the outcome variable was not recorded for that particular treatment.

The interpretation and implications of these results are discussed in detail in the final chapter 5. So, here we limit ourselves to some very simple observations.

First, in comparative terms with the laboratory we certainly find a larger number of significant results, but as in the case of the laboratory such results do not systematically point to a clear direction of what works and what does not work.

Second, in relative terms statistically significant effects are more numerous for the promotional material than for the labels. This has important implications that we address in chapter 5.

Third, the selection of promotional material format is crucial together with process tracing in determining some important results, which we will also come back to in the conclusive chapter.

Fourth, at least for the classification systems the results are fairly clear and, matched with those of the laboratory, can be used to draw some relevant policy implications. We will also come back to this aspect in chapter 5.

Finally, we do not find a strong support for the hypotheses that the highly educated respond in a systematically different way from the rest of the population. We tested for this hypothesis but we find no corroborating results and we do not add many further considerations about this.

Table 7 The summary of the results on behavioural variables

Treatment	WTP	Visualization	At least one Visualization	Complete Visualization
1. Running costs per mile/km	Negative on electric cars			
2. Running costs per 5 years		Positive effect		
3. Lost Saving on fuel as additional costs				Positive Effect
4. Lost Saving on fuel expressed as "you lose" (you save for electric) and the amount				
5. Fuel economy (litres per 100 km/miles)		Positive Effect		
6. Battery Life				
7. German versus Absolute	Not applicable	Not applicable	Not applicable	Not applicable
8. Image of CO2 absolute classification system in promotional material				
9. Explanatory text for the absolute classification system in promotional material		Positive Effect	Positive Effect	
10. Explanatory footnote for running costs in promotional material		Positive Effect		Negative Effect
11. Running costs salience (5 years format) in promotional material		Positive Effect	Positive Effect	
12. Running costs in small format in promotional material	Positive effect (selection of format)	Negative Effect		Negative Effect
13. Web link to full label in promotional material				

Table 8 The summary of the results on noticeability variables

Treatment	Fuel Consumption (except electric)	Fuel efficiency	Running Costs	Electricity Consumption (electric engine only)	Fuel consumption versus standard vehicles (alternative engines only)	Environmental friendliness
1. Running costs per mile/km	• Positive Effect on conventional cars	• Positive Effect on conventional cars	• Positive Effect on conventional cars • Positive effect on electric vehicles	Positive effect on electric		Not applicable
2. Running costs per 5 years			• Negative Effect for Hybrids			Not applicable
3. Lost Saving on fuel expressed as additional costs						Not applicable
4. Lost Saving on fuel expressed as "you lose" ("you save" for electric) and the amount	Positive Effect on conventional cars					Not applicable
5. Fuel economy (litres per 100 km/miles)	Positive effect on standard vehicles					Not applicable
6. Battery Life	Not applicable					Not applicable
7. German/ Absol.	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
8. Image of CO2 absolute classification system in promotional material		Negative effect		Not applicable	Not applicable	• Negative Effect
9. Explanatory text for the absolute classification system in promotional material		• Negative effect (selection of format)	• Negative effect (selection of format)	Not applicable	Not applicable	• Negative effect (selection of format)

Treatment	Fuel Consumption	Fuel efficiency	Running Costs	Electricity Consumption	Fuel consumption versus standard vehicles	Environmental friendliness
10. Explanatory footnote for running costs in promotional material	Not applicable	• Negative effect • Negative effect (selection of format)	Negative effect	Not applicable	Not applicable	• Negative effect • Negative effect (selection of format)
11. Running costs salience (5 years format) in promotional material	Not applicable	• Negative effect • Negative effect (selection of format)	• Negative effect • Negative effect (selection of format)	Not applicable	Not applicable	• Negative effect • Negative effect (selection of format)
12. Running costs in small format in promotional material	Not applicable	• Negative effect	Negative effect	Not applicable	Not applicable	• Negative effect
13. Web link	Not applicable	Negative effect		Not applicable	Not applicable	Negative effect

Table 9 The summary of the results on cognitive variables

Treatment	Score Fuel Consumption (not for electric engines)	Score Running Costs	Score Electricity Consumption (electric engine only)	Score Environmental friendliness	Ranking
1. Running costs per mile/km	• Negative effect (conventional car)			Not applicable	Not applicable
2. Running costs per 5 years				Not applicable	Not applicable
3. Lost Saving on fuel expressed as additional costs			• Positive effect	Not applicable	Not applicable
4. Lost Saving on fuel expressed as "you lose" and the amount				Not applicable	Not applicable
5. Fuel economy (litres per 100 km/miles)				Not applicable	Not applicable
6. Battery Life	Not applicable		• Positive effect	Not applicable	Not applicable
7. German versus Absolute	Not applicable	Not applicable	Not applicable	Not applicable	Absolute is more effective
8. Image of CO2 absolute classification system in promotional material	Not applicable	• Positive effect •	Not applicable	Positive effect	Not applicable
9. Explanatory text for the absolute classification system in promotional material	Not applicable	• Positive effect (selection of format)	Not applicable	Positive effect (selection of format)	Not applicable
10. Explanatory footnote for running costs in promotional material	Not applicable	•	Not applicable	Positive effect (selection of format)	Not applicable

Treatment	Score Fuel Consumption	Score Running Costs	Score Electricity Consumption	Score Environmental friendliness	Ranking
11. Running costs salience (5 years format) in promotional material	Not applicable	• Positive effect	Not applicable	Positive effect Positive effect (selection of format)	Not applicable
12. Running costs in small format in promotional material	Not applicable	• Negative effect (selection of format)	Not applicable	Negative effect	Not applicable
13. Web link etc.,	Not applicable	•	Not applicable	Positive effect	Not applicable

4.5 Explaining the determinants of labels usage

Besides the recovery of the treatment effects from the experimental tasks and the post-treatment questionnaire, we have also analysed the answers provided to the pre-treatment questionnaire by all 8000 respondents before the randomisation started and they were split into sub-samples (see questions Q1 to Q18 in Annex VII). These questions enable us to model the relations existing among the following variables: 1) labels' comprehension; 2) the multiple set of factors considered important in the purchase decision process , other than eco-friendliness and fuel economy; 3) familiarity with labels; 4) trust in labels; 5) usage of label.

It is worth stressing that the answers enabling this kind of analysis were provided by all respondents before being exposed to the treatments embedded into the various variants of labels and promotional material. This means that: a) this analysis does not aim to, and cannot, model and explain the results of the experiments; b) it models answers based only on the respondents previous experience with labels and, thus, not influenced by possible effects (including social desirability) from expositions to the treatments of our experiment.

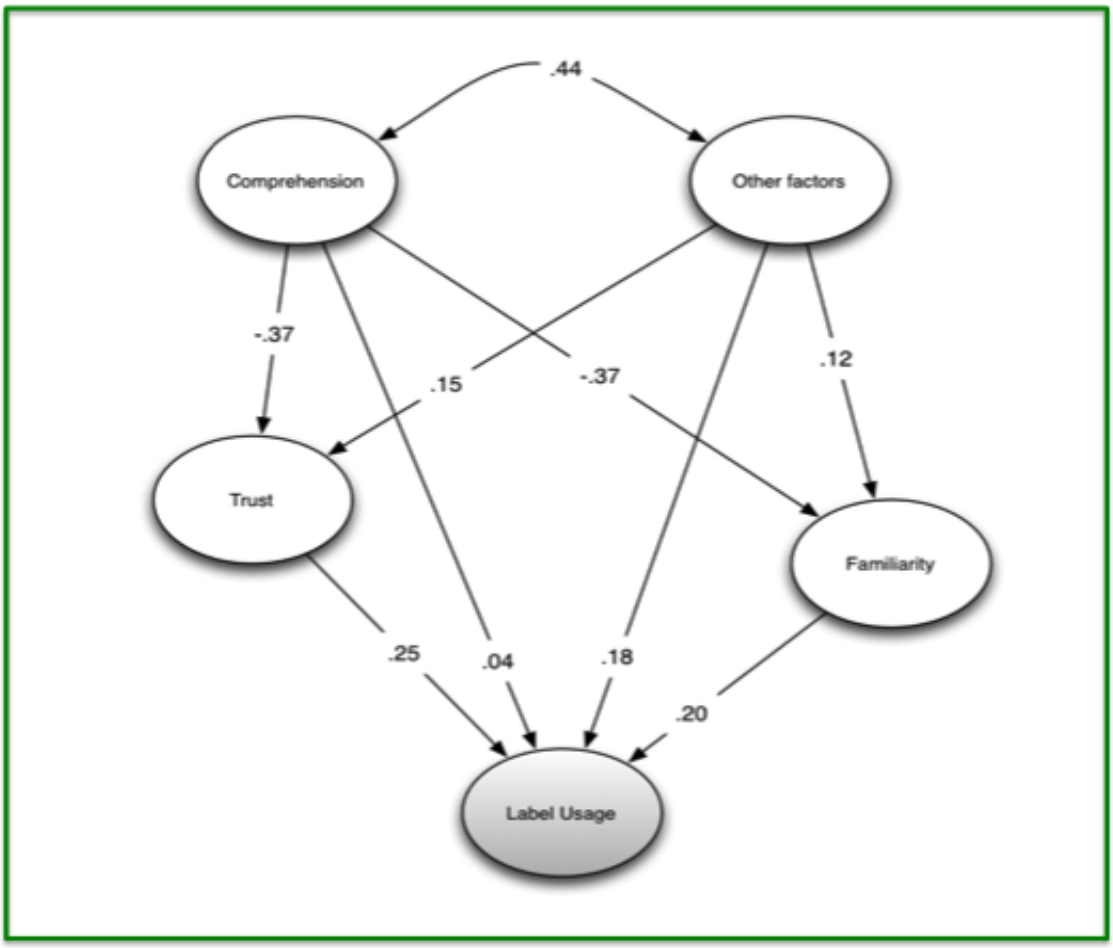
In order to model the relationships existing among the five variables listed above we use the technique called Structural Equation Models (SEM).[21] The result of the general model specified and tested are summarised in Figure 12. The general model was found to be a good fit (X^2= 5698.390, df=408, p <.000). The model shows four 'first order effect' latent variables:

- ***Comprehension (comp)***. The latent variable indicating the level of comprehension of the information contained in eco-labels;
- ***Otherfacts (otherfacts)***. The latent variable indicating the multidimensional nature of car choice. In other words, the importance of other factors in car selection other than the information provided in the label (budget, aesthetic values, etc.).
- ***Trust.*** The latent variable expressing the trust in the information of the labels.

[21] Structural Equation Modelling is a statistical technique that allows researchers to model unobserved variables. This technique is often used to model latent constructs – abstract psychological variables such as "intelligence" or "attitude toward the brand" or " trust" – rather than the manifest variables that are instead used to measure these constructs. By explicitly modelling measurement error, SEM users seek to derive unbiased estimates for the relations between latent constructs. To this end, SEM allows multiple measures to be associated with a single latent construct. A structural equation model implies a structure of the covariance matrix of the measures (hence an alternative name for this field, "analysis of covariance structures"). Once the model's parameters have been estimated, the resulting model-implied covariance matrix can then be compared to an empirical or data-based covariance matrix. If the two matrices are consistent with one another, then the structural equation model can be considered a plausible explanation for relations between the measures. We do not add further technical details here, neither we illustrate the process followed to achieve the final models presented. Such details can be provided upon requesting the authors to do so.

- ***Familiarity.*** The latent variable expressing the familiarity with the labels and the information contained in them.

Figure 12 SEM General Model



We manage to model both the influence that each of these first order effects latent dimensions has on each other and that which they all have on one 'dependent' latent construct (or 'second order effect' latent variables):

- ***Labels Usage (labeluse)***. The latent construct expressing the reasons to use the labels and their information and the intentional or actual usage made of them.

After describing all the constructs in the model, we can now discuss their relationships. SEM modelling produces *co-variances* between latent variables (a two-ways link) and *regression weights* between latent variables (a one way link). Regression weights are presented standardized and can vary from 0 to 1. Observing the paths between latent variables we can derive the following interesting considerations:

1. If we look at the regression weights that affect the use of labels (*labeluse*), we can notice that trust plays an important direct role (.25), as do other factors in car choice (*otherfact*=.18) and familiarity (*familiarity*=.20). The comprehension of labels (*comp*) plays, instead, a small direct role (*comp*=.04).
2. However, the *comprehension of a label affects the usage of labels in an indirect way*. It can be noticed that comprehension has an important relationship with familiarity. The weight is negative indicating that low comprehension affects negatively familiarity, which in turn affect labels usage. In other words, a scarce understanding of labels does not promote their familiarity. Exactly the same happens with trust: low comprehension of labels does not promote trust. Both trust and familiarity are the most important factors influencing the label use.

The implication of the SEM results is that trust and familiarity are important mediators of the impact of label comprehension on labels usage. It should not have escaped the attentive reader that the comprehension of labels is the most important cognitive dimension, which could be impacted through policy nudges such as those we tested both in the laboratory experiment and in the online experiment. The SEM general model, for the overall sample, tells us that comprehension has almost no direct impact on label usage, but all its impact is mediated through familiarity and trust. Therefore, providing the best and most effective nudges will not impact label usage (which in turn is expected to increase eco-friendly purchase and behaviour in using means of transportation) in the short run, but will gradually impact positively familiarity and trust and through them the usage of labels. This means that policy nudges should be provided steadily and can be expected to produce the desired effect as time goes by. In the next three figures we provide the SEM models for three different age groups: a) 18-25 years (age group 1 or young people); b) 25-44 years old (age group 2 or adult); c) 45-65 (age group 3 or middle aged to senior)

Figure 13 SEM Model for young people (18-24)

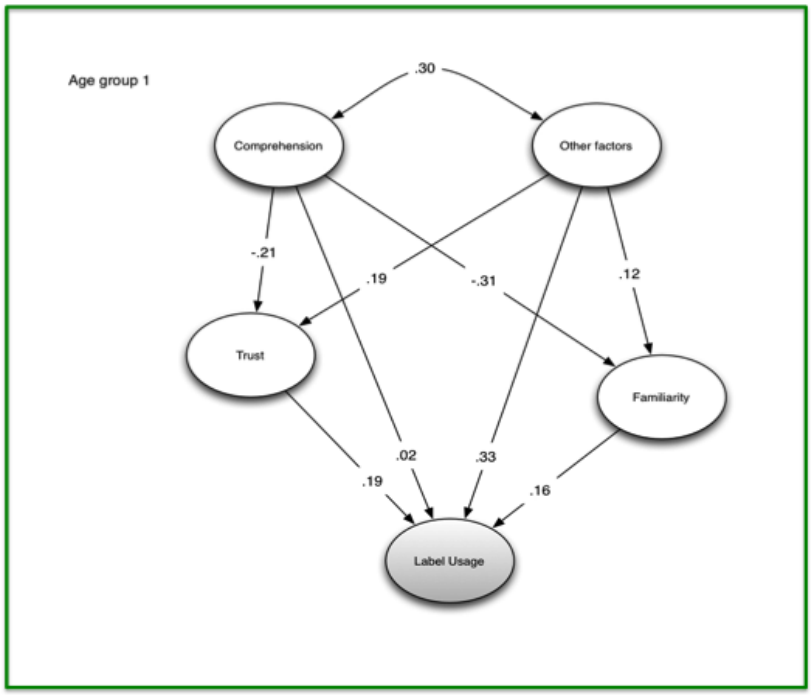


Figure 14 SEM Model for adults (25-44)

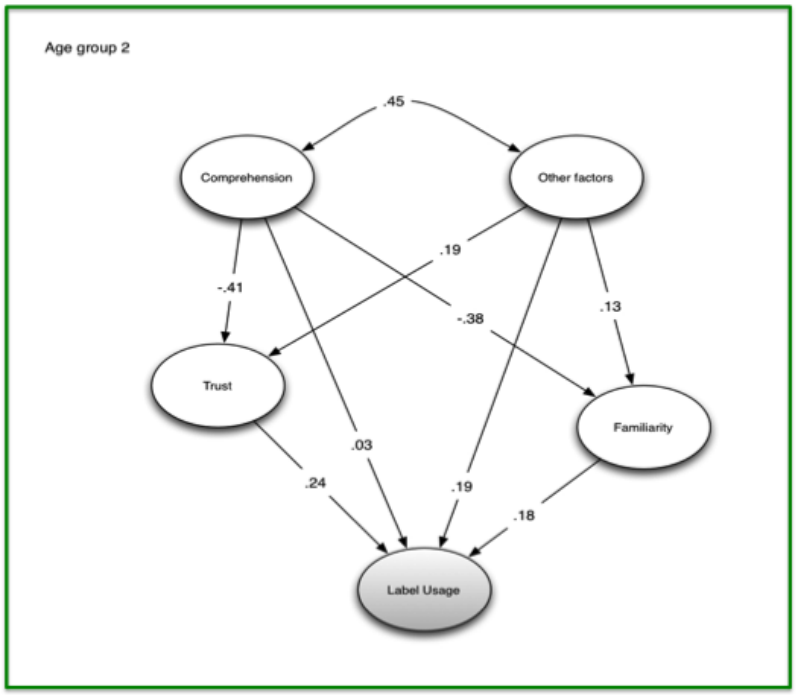
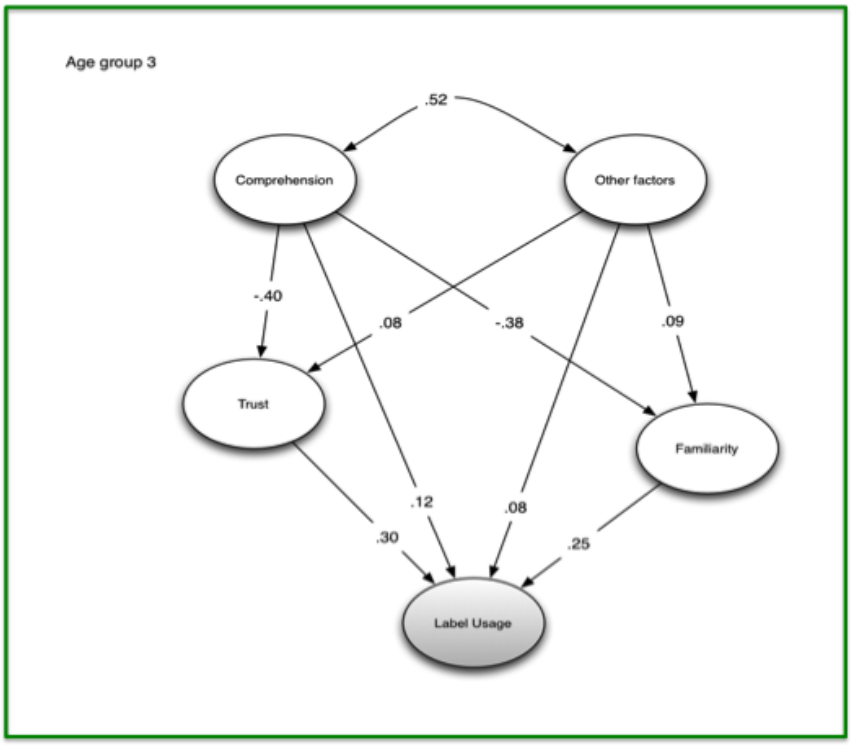


Figure 15 SEM Model for middle-aged and seniors (45-65)



Comparing the model for young people (Figure 13) with the general one (Figure 12), we notice that:

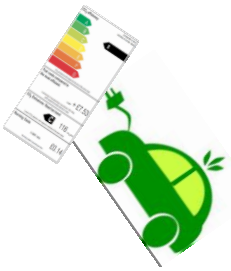
1. The latent construct 'other factors' has a larger influence for younger respondents in their use of the label. This probably means that the choice is largely shaped by factors such as budget, aesthetic, performance, etc.
2. Although the relationship remains the same, the weights of comprehension are smaller on both familiarity and trust indicating that, for younger citizens, the understanding of the information in the labels affects less trust and familiarity. The 'other factors' are also here slightly more important for trust as indicated by a larger weight.

Comparing the general model with the one of age group 2 (25-45 years old; Figure 14), we see that in this group the importance of comprehension for trust is higher. This group represents a mix between the model of younger citizens (age group 1) and the general model. The latent construct 'other factors' does have a larger role in trust but about the same in familiarity and label usage. The higher age suggests a larger impact of comprehension on trust and familiarity compared to younger respondents but it shares with them the influence of other factors.

Finally, when we look at the SEM model for middle-aged and seniors (Figure 15) it is interesting to notice that this model has some significant differences compared to the general model and the models of the two previous younger age groups. For older citizens, trust, comprehension and familiarity have a larger influence compared to younger respondents while 'other factors' are less important. For this age group, understanding the information is a crucial element in using labels, through the direct and indirect ('mediation') ways in which comprehension of labels affect trust, familiarity and label usage itself. In other words, for this type of citizens, understanding labels, being familiar with and trusting them is the path towards their usage. Instead, 'other factors' that tend to be important for the youngest and the middle-age groups are less important in this case.

In conclusion, we would like to propose the following interpretations of the SEM analysis:

- *Designing labels ,which are easier to understand, does play an important role but this role is indirect and mediated by trust and familiarity. Increasing comprehension of labels leads to higher trust and familiarity, that in turn leads to increased labels usage;*
- *On the other hand, the importance of greater comprehension of labels, that can be attained through policy nudges, seems to increase with age whilst that of other factors decrease and, probably, also as intuitive cognitive capacities decrease.*



5 Conclusionsand Recommendations

In the three core chapters (2, 3, 4) we mostly reported the findings in neutral way, simply stating and re-phrasing what the processing of the data tells us in the form of descriptive statistics, multivariate statistical analyses, and of econometric analysis (i.e. regressions to recover the treatments effects). In this chapter we recall some of these 'objective' findings but we also add some interpretations and inferences that help us extract the policy relevant implications and then formulate a few basic and simple but promising policy recommendations, both on substance and method.

In developing this kind of final analysis the narrative will transparently alert the reader as to when we are interpreting, and inferring, from the results and, thus, we believe there is no need to add a disclaimer whenever we venture into reading in the results something that the numbers in the technical results tables presented in the Technical Compendium do not directly and strictly tell us.

We do this not on the basis of imaginative thinking but rather triangulating the results of the three empirical components among each other and with the relevant literature.

In particular some of the considerations that follow are inspired by, and cite, the insights from behavioural economics that we presented in § 1.2 (see Table 1, p.26; Figure 2, p.28; and text around them). So, we take for granted and do not explain again concepts such as System 1 and System 2, hot and cool cognition, heuristics, and biases.

We also use the distinction between a 'main effect' and a 'full factorial' design and its implications in terms of what we could and did test (and of what we could not and we did not test) without entering into further explanation to those already provided (at the end of § 1.2 and also at the end of § 3.1).

5.1 Disclaimer and implications for future studies

This disclaimer concerns the results of the two experiments and helps us suggest ways to empirically assess behavioural effects of labels and promotional material in the laboratory-controlled environment in the future. As we stated in the introduction of this report our study is the first to test car eco-labels and promotional material through randomised control trials with double randomisation, and in ten European countries (the online experiment). As such it has produced a wealth of new empirical evidence that can shape future studies further exploring some of our findings and the interpretative hypotheses we present in this chapter.

Therefore, we do not aim to provide in this chapter the conclusive words on what works and what does not work. Our aim is to present interesting findings and interpretations on which other researchers and policy makers can build, improve, and so advance our behavioural understanding of how to use eco-labels aimed at 'nudging' more sensible consumers' choices that can benefit European economies and societies.

Having said this, we must also stress that the results we have presented in chapters 3 and 4, and commented in this chapter, have been produced through a robust and internally valid design with a fair degree of external validity. The effects we found being significant or not significant are robust across different specifications of the regression techniques used, and hold when controlling for age, gender, and country (this item only for the online experiment). In the domain of what they claim, they can hardly be challenged. As explained, what they tried to demonstrate is whether or not the various experimental conditions we tested (i.e. information elements and/or their layout) have a causal effect in a counterfactual sense, which means always comparing a treated group with a control group.

Nonetheless there are a few results that may appear counterintuitive from a substantive perspective and may run counter to what one would have expected using only common sense (i.e. the impact on the web link in promotional material, discussed above). We cannot rule out that these results may be due to two technical aspects that were beyond our control and that we could not implement differently.

The first concerns the fact that we could not freely design the nudges in such a way as to make the difference between the treatment and the control sharper and/or as to follow more closely the biases and corresponding de-biasing mechanisms one could extract from the behavioural economics literature. As we explained in chapter 1, we were constrained in the sense that the treatments tested could not deviate too much from general regulatory prescriptions and from the labels and promotional material that are already

currently in use. As a result, the difference between control and treatments is not as sharp as it would have been if we used, for instance, a very different treatment such as the one exemplified in Figure 3(p. 29).

The second constraint consists in the fact that with the large number of treatments we were asked to test we could not implement a full factorial design but rather a main factor design. The former would have allowed capturing interaction effects between different information elements and, in practice, to test labels as holistic treatments instead of information elements one at a time. We recall, in fact, that we did not test labels or promotional material as such, but we tested the various different elements contained in them. A full factorial design where interactions are captured may have produced different results for some of the elements (but certainly not for all)

On the first constraint we do not think that much can be changed in the future, since testing nudges that radically deviate from what may be feasible to implement, from a policy perspective, may produce better results that would not be policy compatible.

On the contrary, with regard to the second constraint, future policy applied studies should try and implement the principle ***'less is more'***. For instance if one considers only two factors such as "eco-friendliness" and "fuel economy", each containing only three values, and only for conventional car engines, then it would have a simple 3*3 full factorial design. According to the rule of thumb explained above (3.1 and Technical Compendium) this could be implemented with around 300 subjects (in a clean between-subject design without need for participants to be exposed to more than one experimental condition). Under this ideal scenario then one would actually test different labels as holistic treatments and may reach more conclusive results.

5.2 Preliminary survey

Awareness about eco-friendly behaviour is considerable …

Our survey confirms that Europeans are aware and not indifferent to the environmental problems caused by the pollution produced by vehicles, and there are several indications of this.

First, whereas currently only about 1.9% of the sample possess either an hybrid (1.8%) or an electric (0.1%) vehicle, as many as almost 33% of respondents say they will buy electric or hybrid as their next car.

Second, more than 50% of the respondents think that more than 40% of the greenhouse effect is attributable to car pollution.

Third, while most respondents do not show much faith in the eco-friendly behaviour of others, nonetheless a clear majority of them is aware that their behaviour has an impact on the environment and thinks that they can make a change with their actions.

… but there is clear 'attitudes-action gap'

On the other hand, our results also confirm the gap observed in literature between self-reported attitudes/intentions and actual behaviours. There are several indications of this.

First, environmental concerns come after 10 other main attributes (price, safety, performance, etc.) in terms of importance in influencing car purchase decisions.

Second, after showing awareness about the environmental impact of car usage, when asked a question on preferences clearly related to their actual behaviour, the majority of respondents revealed a preference for their own private car as compared to alternative means of transportation for daily commuting.

Third, the preliminary survey confirms that consumers first select a class of vehicles, and only when they narrow down to choosing a model they may then take into consideration eco-friendly parameters.

We find both moderate familiarity with the eco-labels and high credibility for them

The set of questions on current existing environmental labels provides a mixed picture with both positive and negative signals. More than half of respondents report not being very familiar with labels, 40% disagree with the statement that they are easily recognisable, and 44.5% agree that car labels are unfamiliar to them. Many also misunderstand environmental labels as symbolising product reliability. On the other hand, perceived credibility is fairly high, although many think that the information contained in the label is not sufficient.

Results of the split ballot: relative system more effective?

The split ballot included at the end of the preliminary survey fulfilled only an exploratory purpose to help decide which classification system should be tested experimentally in the laboratory. Its scope was very limited in that, due to time constraints (split ballot came after 36 questions in the survey), only one car was used and this may have influenced the relative effectiveness of the different classification systems on the dependent variables measuring noticeability (Q37a, Q37b, Q37c, Q37d). With this caveat in mind, the results of the analysis show that:

- The relative classification system, compared to the other three (absolute, combined, and German) had the highest impact in terms of positive evaluation of environmental friendliness and fuel efficiency (response variables) made by the respondents on the car they had seen (a VW Polo);
- For all response variables, the German classification system induced less positive evaluations;
- For all response variables, the combined classification system had little or no impact;
- The effects were, however, statistically significant but small;

5.3 Laboratory experiment

Some interesting but non-systematic findings

In the laboratory experiment we found a number of treatments that have positive and statistically significant effects and, thus, the first screening has not been inconclusive. On the other hand, it is clear that that the results are not consistent across treatments, measures, and engine types, but somewhat scattered across them. In other words, we cannot draw clear-cut conclusions on the most effective treatments, especially since pairwise comparisons of effective treatments do not yield conclusive results, except in the case of noticeability (response variable) and classification systems (treatments), where the Relative System solution appears slightly more effective than the Absolute one and clearly more effective than the Combined one.

General interpretation: nudging System 1 or System 2?

In the case of eco-labels there is no possibility of using the very effective nudge represented by changing defaults. For reasons explained earlier the use of nudges acting directly on emotion (as for instance the fictitious example in Figure 3, p.29) was unfeasible.

Therefore, the variants of information elements and/or their layout that we tested fall into the category of nudges that, by better framing existing information elements and/or by including clear and easy to understand new elements, aim to activate reflective and cool cognition in consumers.

The interpretative hypothesis we propose for the non-systematic, and scattered results we found, is that our nudges in the laboratory experiment (the reasoning also applies to some extent to the online experiment) were still too overloaded with information and had mixed effects. They obviously did not affect directly System 1 emotion but most of them were also too complex to fully activate the more reflective and cool cognition of System 2.

If we consider conventional cars we have a few noticeability effects (mostly of the more graphic elements) but almost null effect on cognitive recall of information. System 1 is not triggered enough and the information contained is still too much to be recalled by a not fully activated System 2. In order to be processed cognitively the information contained in the visual stimuli requires a slow, controlled, and conscious process (Samson & Voyer, 2012). If the treatments, that would go into a label, require cognitive efforts as intense as other forms of information provision, then they are missing their goal. Lacking full

activation, then it is perfectly explainable why we do not see any effects on the behavioural choices (as measured by willingness to pay and related variables)

On the other hand, cognitive measures with respect to the fuel economy show somewhat better results for the impact of electricity consumption, running costs per mile and per 5 years in the case of the electric cars. In this case we can propose the following interpretation in line with the main hypothesis presented here. The information is much starker and might benefit from the subconscious assumptions about the electric car, thus, producing an effect where cognitive processing is less demanding. These same treatments for electric also show positive effects on behavioural measures, thus reinforcing our reasoning. The labels work both on behavioural and cognitive response variables when they are easier to process and more likely to help intuitive thinking. If this interpretations is correct than policy makers face the dilemma between including in the labels all the information required by statutory constraints and having less charged labels that are effective as nudges of eco-friendly and fuel-efficient consumers choices.

Fuel economy nudges seem to work better

The second interpretative hypothesis we propose is that nudges concerning the fuel economy both work better than those concerned with more strictly defined environmental issues (i.e. CO2 emissions). We find the following effective elements (on some but not all of the regressed response variables):

- For conventional engines cars:
  - Information on fuel lost savings
  - Running costs per mile
  - Running costs per 5 years
- For electric cars:
  - Information on electricity consumption
  - Running costs per mile
  - Running costs per month
  - Running costs per 5 years
- For hybrid cars:
  - Combined information on consumption
  - Separate information on consumption
  - Running costs per month
  - Running costs per 5 years

In addition battery range and running costs per 5 years in the case of electric have a significant effect also on Change in Willingness To Pay (CWTP). In addition, the treatments not related to the fuel economy, like the graphic layout or the classification systems,

seem to impact more on behavioural choices related to the fuel economy than on those related, for instance, to CO2 emissions.

This interpretation of the laboratory results is fully in line with the scientific literature reviewed and with the findings of our preliminary survey. In a report delivered in May 2012 for the Low Carbon Vehicle Partnership (Lane et al., 2012), based on a survey of consumers, great emphasis is placed on running costs both prescriptively and empirically. On the prescriptive dimension the authors report that the increased emphasis on running costs and financial information is part of a strategy adopted in the UK to nudge consumers toward buying more eco-friendly (Lane, et al., 2012, p. 15). Indeed, both our preliminary survey and the laboratory experiment support (though in different way) the idea that nudges based on economic aspects can be more effective than nudges based exclusively on environmental considerations. Empirically the authors, of the cited report, show that the respondents in their survey were particularly impressed by the information on running costs and extensively report statements, made by participants in focus groups, on the importance to them of running costs (Lane, et al., 2012). Indeed, this finding is not too dissimilar from those of our preliminary survey. Running costs are considered an important attribute by 69.5% of our sample, and this result is confirmed across both the socio-demographics variables and across countries.

5.4 Online experiment

Larger number of significant effects but still fragmented (especially for labels)

The fact that the results of the laboratory experiment were not consistent across different reference point could either be interpreted as due to the comprehensive measurement strategy (inclusion of a wide array of response variables) or as deriving from the conditions included and tested through the labels. In order to test this hypothesis in the online experiment we have tried to use more homogenous response variables (increasing space for behavioural ones) and we have simplified and reduced the number of conditions tested. As we anticipated in the synoptic summary, we find a larger number of significant results in the online experiment as compared to the laboratory one. However, despite the changes made the results are still fragmented and non systematic across treatments of the same dimension, type of engines, etc. Especially for the labels we can conclude that this fragmentation is not caused by the measurement strategy but is rather related to the nature of the nudges we tested.

We can draw a conclusive assessment of classification systems

One question, however, where we can be more conclusive using the results of all three empirical components is that of the classification system. There are no doubts that the choice should be restricted between the Absolute and the Relative classification systems. The relative system came out stronger in the preliminary survey split ballot and in the laboratory experiment. In the online experiment, the comparison between the absolute classification system and the German classification system (a variant of relative classification) clearly showed the former to be better than the latter with respect to cognitive processing. Respondents exposed to the absolute systems ranked the car they saw in terms of CO2 emission more correctly than the respondents who were shown the German classification system.

In strictly statistical terms we cannot conclude what is the most effective between the relative and the absolute systems. In the Laboratory experiment the pairwise comparison between Absolute and Relative, where they are both significant, is not conclusive, and we gave a slight preference to the Relative system as its effectiveness was recorded on a larger array of measurement scales. In the online experiment the Relative was not tested and so we cannot compare it with the Absolute.

The comparison can be made, however, taking in consideration other elements such as what we know about the purchase process. In a context where consumers first decide a class of car and second analyse the kind of information contained in eco-labels, then the most pragmatic solution would be to give them a comparison with other cars in that class

(relative classification system). On the other hand, even at first sight it is clear that the Absolute classification system is the simplest and the easiest to process and understand. So, in a policy that (as we suggest later) in the short term should aim at increasing familiarity, trust, and comprehension it seems more appropriate to use the Absolute system even though this might not be sufficient to have consumers change their preference of class. As familiarity, trust, and capacity to process the information, in the label, increases over time then in a subsequent period the Relative system may be introduced or combined.

Again on the same issue: nudging System 1 or System 2?

With the exception of a counterintuitive effect with electric car (that we will discuss afterwards), we do not find significant results for Willingness to Pay (WTP) under classical experimental conditions. Here again this can be explained by the information overload and the difficulty to fully process all items when requested to evaluate cars and decide at which price to buy them.

It is interesting to notice how the results change when the Multiple Price List task (from which WTP is elicited) is performed for the promotional material in the format selected by the respondents, and with process tracing. The respondents that have to select the format, and use the mouse to visualize, are willing to pay more, increasing the impact of eco-friendliness and/or appreciating the format provided. Here what happens, in our interpretation, is the following: a) having to deal with mouse-over in a context in which they have to choose the format, respondents make an extra effort to read and cognitively process the information; b) this more careful examination of the label may then shape their evaluation and the choice the make. The choice requested, and the process tracing, activate participants and the result is that they use their System 2. This result is very important because it indirectly shows that at least some bias can be de-biased through opportune choice of architecture.

Promotional material seems more effective: simplicity, variation, or familiarity?

It is pretty evident from the synoptic tables reported in § 4.4 that we find considerably more statistically significant and positive results for promotional material compared to the labels. Moreover, if we take very similar response variables this is even starker when comparing the promotional material of the online experiment with the labels of the Laboratory experiment. In our view there are three possible interpretations that are not alternative but rather reinforce each other.

First, promotional material visual stimuli are simpler than their label counterparts and, thus, may be easier to process and use, even without a full activation of System 2.

Second, within promotional material there is more variation both within treatments and between them and the control group. Even a quick slideshow presentation, with a small selection of labels and promotional material images, shows a striking similarity across different labels but clear-cut differences across different versions of the promotional material. The labels are constrained by a common structure that makes treatments more similar to each other and, especially, not so different from the control group.

Last but not least, it is quite possible that promotional material may represent for consumers a more familiar form in a purchase context than the labels. This is an important element that can be related to the issue of familiarity, trust, comprehension, and usage of labels that we have seen in the results of the Structural Equation Models exercise presented in § 4.5. We come back to this aspect in the paragraph on general conclusions and policy recommendations

The effectiveness of fuel economy nudges is confirmed with some distinctions

If we triangulate the results of the Laboratory experiment, with those of the online experiment, we can safely conclude that the most systematic results we have found concern treatment belonging to the broadly defined domain of the fuel economy. In the laboratory we found support both for running costs and for lost saving on fuel. In the online experiment the support is stronger for running cost (in either form). Lost saving on fuel does not perform too well and the formulation 'you lose' followed by the amount does not work. Hence, if we put together laboratory and online experiment, we find the most systematic support for the usage of running costs.

Potentially counter-intuitive results: running costs reducing WTP for electric car

Having stated the above about running costs, we must, however, also point out that even in this domain there can be unexpected results. A reasonable expectation would be that information on C02 emissions and on running costs would increase the WTP for eco-friendly cars consuming less fuel or decrease it for eco-unfriendly car consuming more fuel. Yet, we are confronted with the finding that running costs decrease the willingness to pay for electric car. This result could hardly be interpreted unless we consider this as case of a bias concerning System 1. The bias we have in mind is the so called 'valence effect' that we can simplify as 'wishful' thinking. Respondent may have an overly optimistic expectation about the running costs of an electric car, they may assume for instance that they are close to zero, but when confronted with the actual figures they remain frustrated and as a result, their willingness to pay decreases. This result shows once more the possibility that nudges may have counterintuitive effects and should be cautiously designed. Although needing further empirical scrutiny, a similar result suggests the possibility of subsidizing eco-friendly choices such as electric cars.

5.5 Conclusions and recommendation

Increase familiarity and trust with easy to understand labels

The best way to start this final paragraph is to triangulate the results of the experiment with those of a model analysing the relationships among the factors that explain actual usage of labels (presented in § 4.5). The latter clearly tell us that at the general population level it is not realistic, in the short run, to influence the behaviour through the use of eco-labels only by providing better nudges that are easier to process and understand. The direct effect of labels comprehension on labels usage is very limited, whereas he effect of familiarity and trust on labels usage is stronger. On the other hand, comprehension affects both familiarity and trusts that are mediators of its effect on label usage.

We can, thus, re-read the results of the two experiments in light of this general explanation of label usage and advance the following interpretative hypothesis.

We do not find systematic effects for the nudges tested, not only because they do not activate efficiently cool and reflective cognitive processing, but also because respondents are not so familiar with labels. Low familiarity with labels was also found in the preliminary survey. As familiarity and trust increase with time then an easier and more intuitive processing of the labels might increase their effects.

On the other hand, improving the comprehension is also important since the explanatory models of labels' usage run for different age groups show that comprehension becomes more important as the age of individual increases. Since car purchase decisions are concentrated in the adult and middle-aged population working on the comprehension of labels is also important.

Hence, the most general policy recommendations is that the Commission and Member States should use the most effective and easiest to understand label in order to gradually improve comprehension and, thus, affect familiarity and trust and eventually impact label usage.

Absolutely exploit the promotional material channel

In order to reinforce and speed up the process of familiarisation, it is of the uttermost importance that the promotional material channel is exploited for conveying to consumers the correct nudges. Promotional material is more familiar for consumers, is less constrained by a fixed structure, and we have shown it to have quite important effects on a wide array of the response variables used.

Labels should be based on:

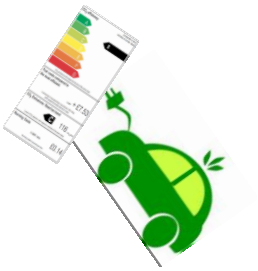
- Vertical layout;
- Absolute classification system;
- Running cost per mile/km to be rendered with a graphic colour based rating that is placed on the side of the actual amount;

Promotional material should be based on:

The treatment we used and that is summarised in the figure below

Figure 16Recommended promotional material layout

- Graphic element for car classification using absolute classification;
- An additional textual explanation on CO2 emissions;
- Running cost salience (possibly to be rendered with coloured based rating as proposed for labels) and accompanied by a footnote explaining the unit of measurement.

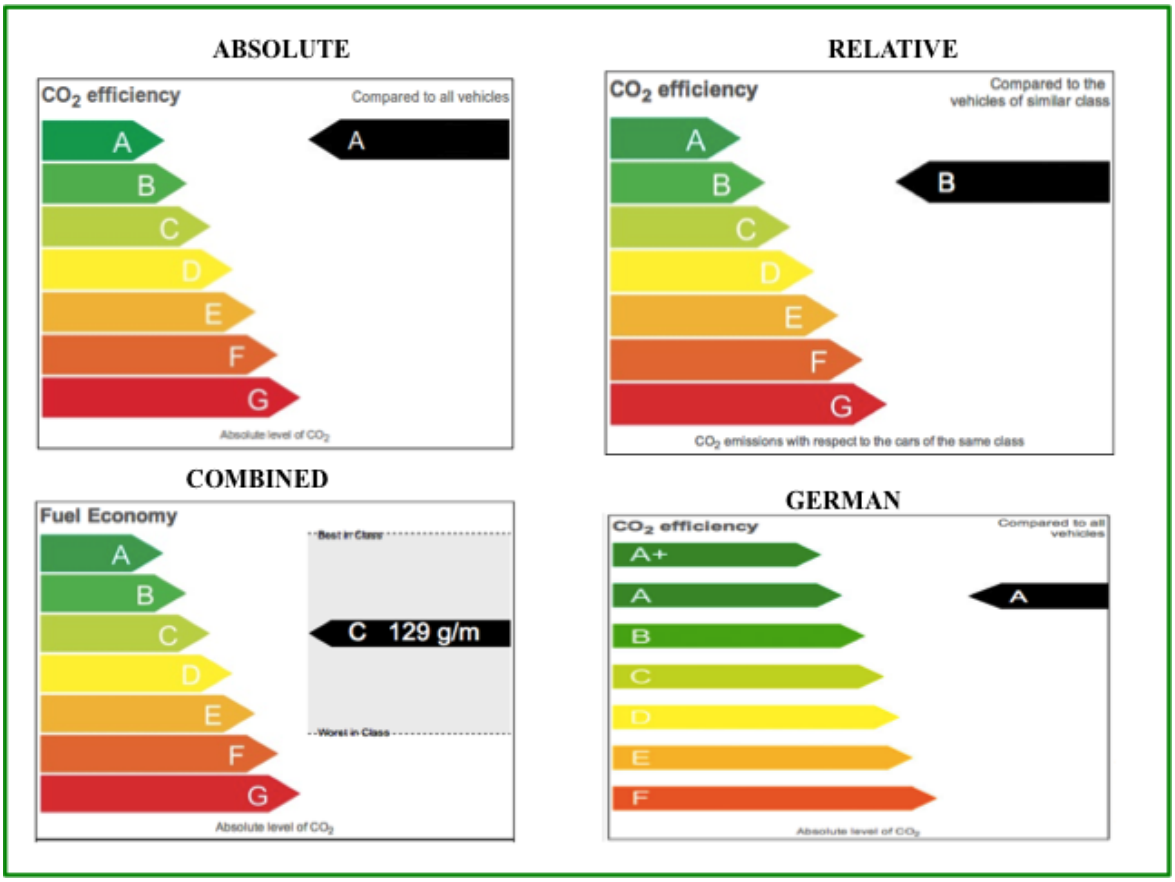


6 Appendix: treatments and visual stimuli

6.1 CO2 emission classification systems for split ballot

Below we report the image of the CO2 classification systems used for the split ballot of the preliminary survey.

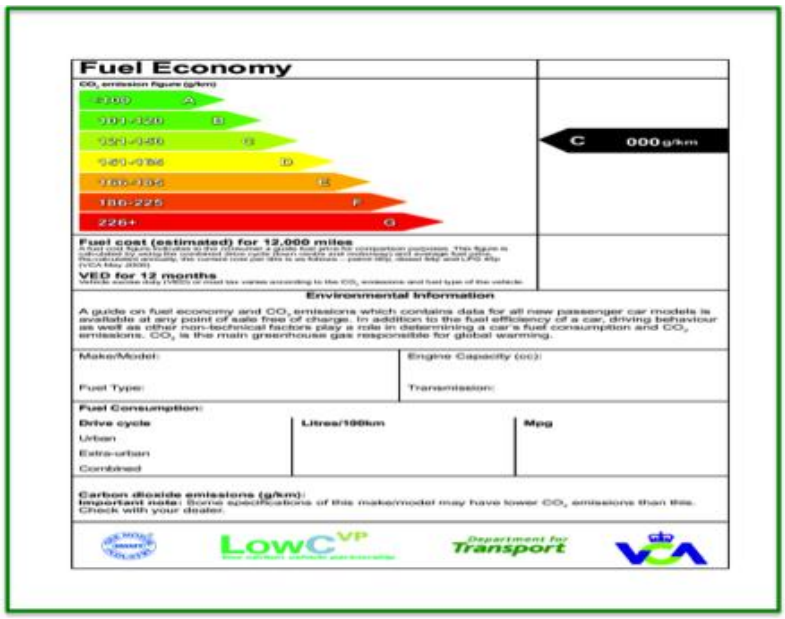
Figure 17 Classification systems used for CO2 emissions



6.2 Laboratory experiment

1. **Control condition (placebo)**: standard label currently in use in the United Kingdom:

Figure 18 Control condition: UK standard label

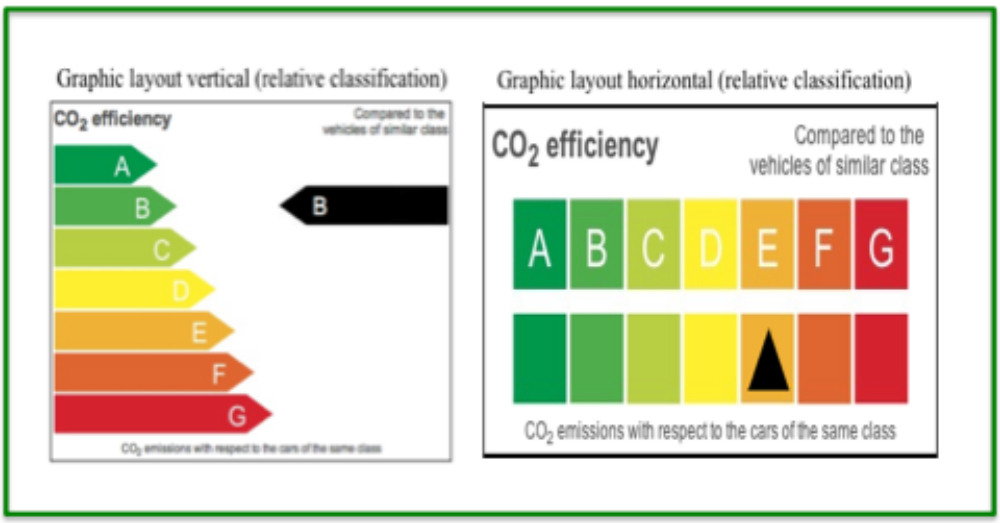


2. **Graphic layout of the classification system**:

   a. Vertical;

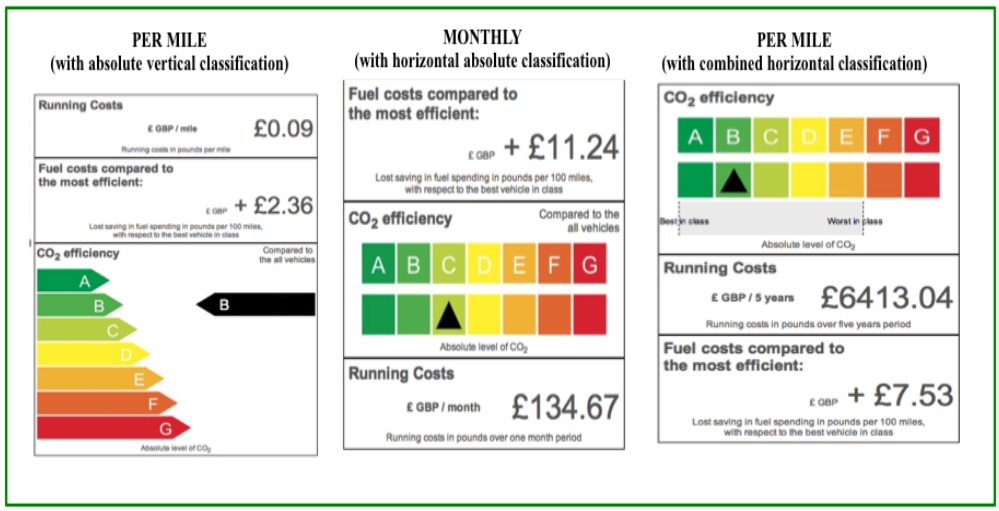
   b. Horizontal;

Figure 19 Graphic layout (example based on relative classification)



3. **Alternative classification systems in terms of CO2 emissions**: (see Figure 17, page 98)
    a. Absolute, car compared to all cars;
    b. Relative, car compared to cars of similar class;
    c. Combined, combination of the two;
4. **Additional Information: running costs**(three possible levels, see figure below)[22]:
    a. Per mile[23];
    b. Monthly;
    c. Per 5 years;

Figure 20 Exemplification of running costs levels



5. **Additional information: Lost savings on fuel spending.** Present or absent (see in figure above). The information is always formulated in terms of lost savings per mile compared to best vehicle in class and is independent of the classification system used and of the running costs format;

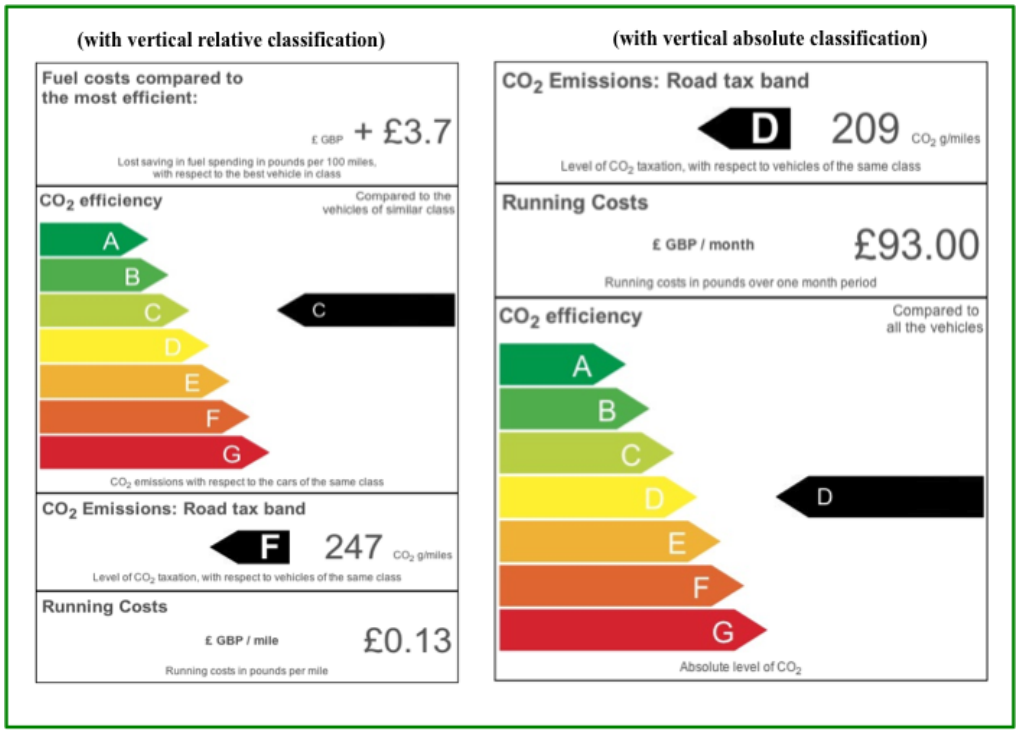
6. **Additional information (for conventional cars)**:**Level of CO2 taxation.** Present or absent. The information is always formulated in terms of level of taxation

[22] The three labels reported in the figure are only exemplificative of the way the three running costs levels above appeared in the various labels shown to respondents in conjunction with the cars randomly assigned to them in relation with the simulated purchase task. This comment on the exemplificative and non-exhaustive nature of the figures applies for all other experimental conditions and will not be repeated.

[23] The laboratory was held in London. In Online experiment this condition is Euro/Km (or other currency where necessary), except for the samples of participants from UK.

compared to best vehicle in class and is independent of the classification system used and of the running costs format (see Figure 21)

Figure 21 Exemplification of CO2 taxation (conventional cars only)



7. **Additional information for electric cars**:
    a. Tailpipe and upstream emissions(two possible levels, see Figure 22):
        i. 'Tailpipe only'
        ii. 'Tailpipe and Upstream' (as two separate items);
    b. Consumption (two possible levels, see Figure 23):
        iii. Range in distance covered with fully charged battery;
        iv. Electricity consumption;
8. **Additional information for hybrid cars**:
    a. Tailpipe and upstream emissions(two possible levels, see Figure 23):
        i. 'Tailpipe only'
        ii. 'Tailpipe and Upstream (as two separate items);
    b. Consumption (two possible levels, see Figure 23):
        i. Two separate figures for fuel consumption and other source of consumptions (Separate);
        ii. One synthetic indicator for the two (Combined)

Figure 22 Exemplification of additional information: electric cars

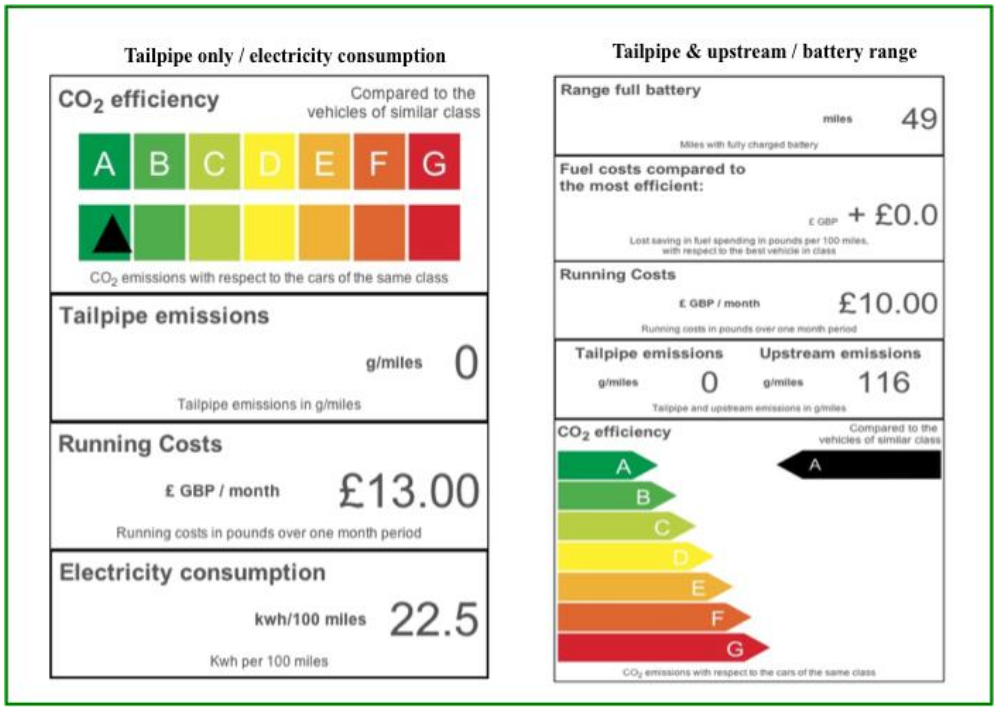
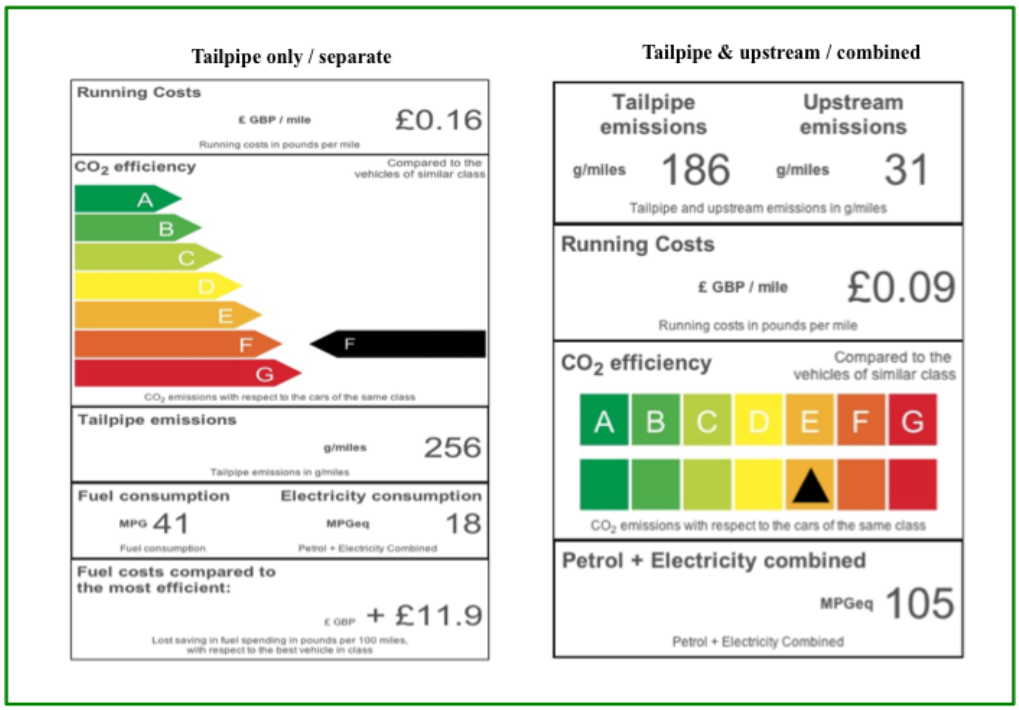


Figure 23 Exemplification of additional information: hybrid cars



6.3 Online experiment

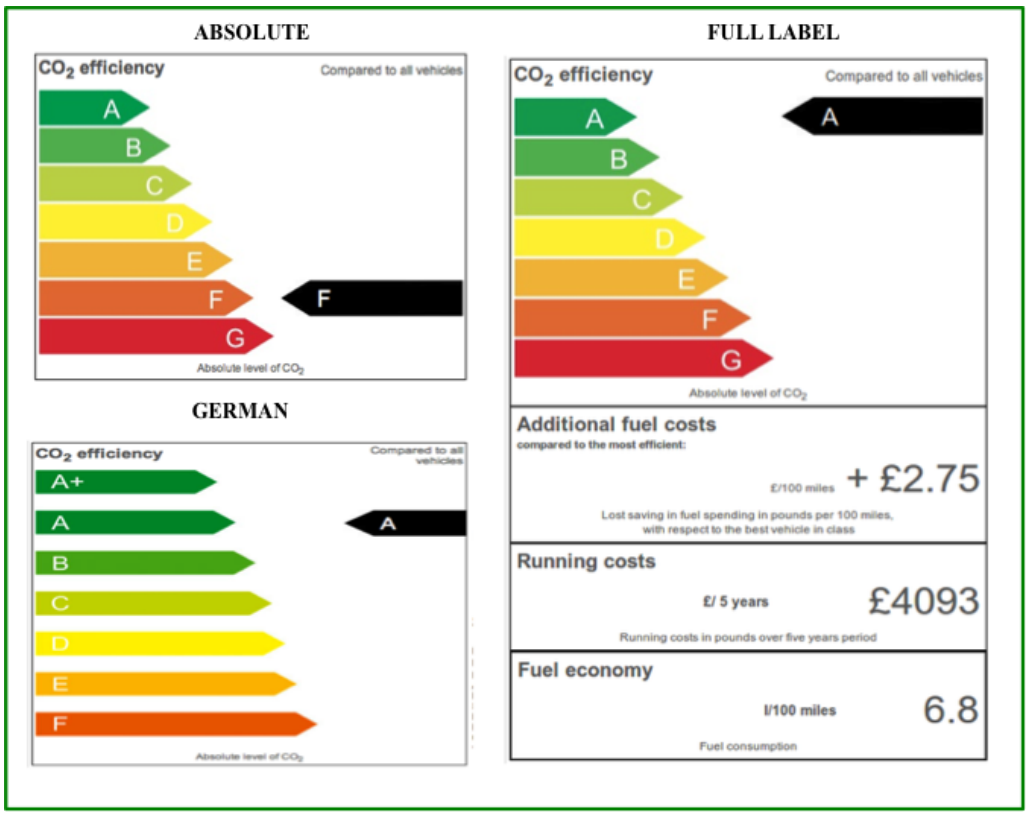
For obvious reasons of space and of language we report only the visual stimuli used for the UK, but we produced version in all other nine languages.

Task 1: Re-test of labels (in simplified format compared to laboratory experiment)

► **Sub-Task 1.1: CO2classification systems only for conventional cars (N=800).**

- German classification system with simplified label;
- Absoluteclassification system with simplified label;
- Full Label (with either German or Absolute classification

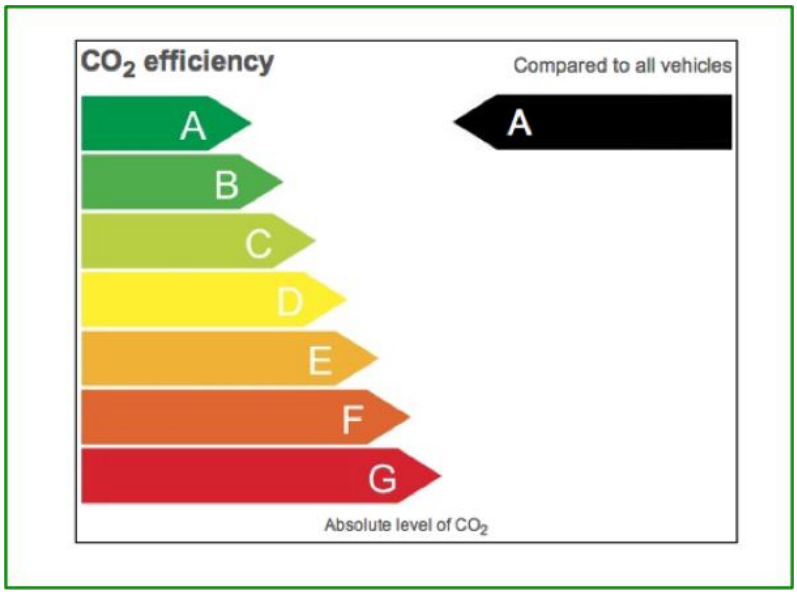
Figure 24 Absolute vs. German simplified label vs. full label (exemplification)[24]



[24] Based on UK label (absolute rating).

- **Sub-task 1.2: further test of labels in simplified format (N=3200)**.
  - **Control condition** (standard label with vertical CO2 absolute classification)

Figure 25 Control condition (re-test of labels, online experiment)



  - **Running Costs (conventional and hybrid)**:'cost per mile/km' vs. 'cost per 5 years'

Figure 26 Treatments for conventional and hybrid (online experiment)

- **Lost savings on fuel (conventional and hybrid)**[25]**:**as additional costs vs. text "you lose" followed by the additional costs (see Figure 26);
- **Fuel economy (conventional and hybrid):**litres per km or miles per gallon depending on the country (see Figure 26);

- **Running costs (electric cars)**: 'cost per mile/km' vs. 'cost per 5 years' (see Figure 27);

- **Savings on fuel (electric car):** expressed with the text "you save" and the money value, savings calculated comparing with the most efficient conventional engine car in the same size class(see Figure 27);

- **Battery life:** miles/km with fully charged battery (see Figure 27);

Figure 27 Treatments for electric (re-test of labels, online experiment)

25 As compared to the most efficient model in a class.

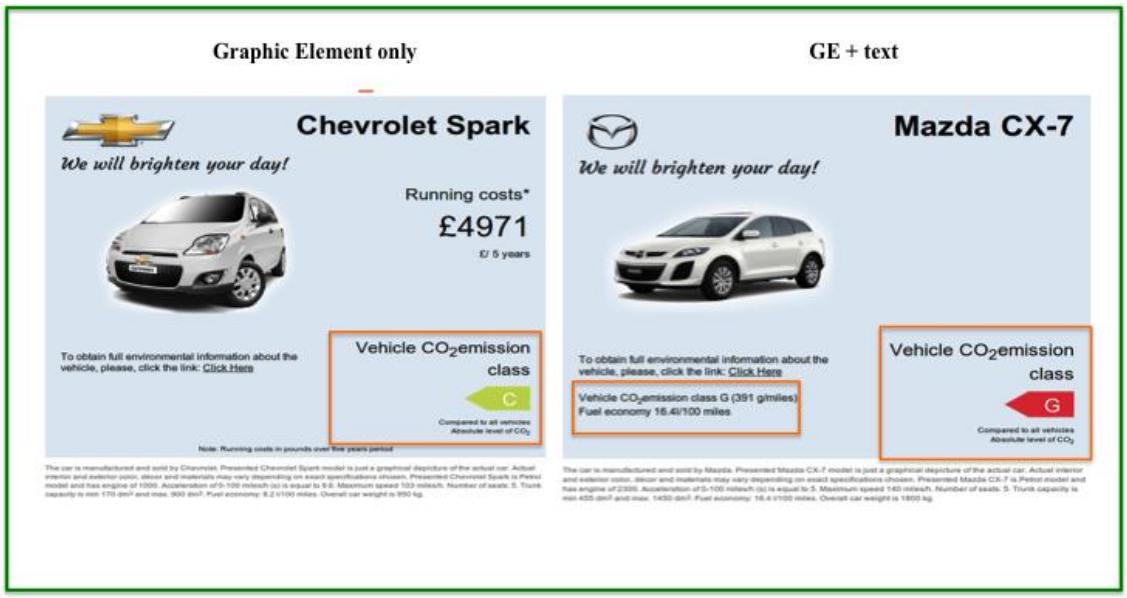
Task 2: Promotional material.

- **Control condition** (format in which they are currently available on the market);

Figure 28 Control condition (promotional material, online experiment)

- **General format** (see Figure 29):
  - **Graphic element**: information about CO2 emissions in graphic format;
  - **Graphic element and text**:information about CO2 emissions in both graphic and text format;

Figure 29 General Format (promotional material, online experiment)



- **Additional Element**(see Figure 30 - Figure 32below)**:**
  - **Class**: containing only the CO2 emission class;
  - **RC small**: containing the CO2 emission class plus a small text indicating running costs;
  - **RC salience**: containing the CO2 emission class plus a larger running cost element;

Figure 30 Additional element (promotional material, online experiment) 1

Figure 31 Additional element (promotional material, online experiment) 2

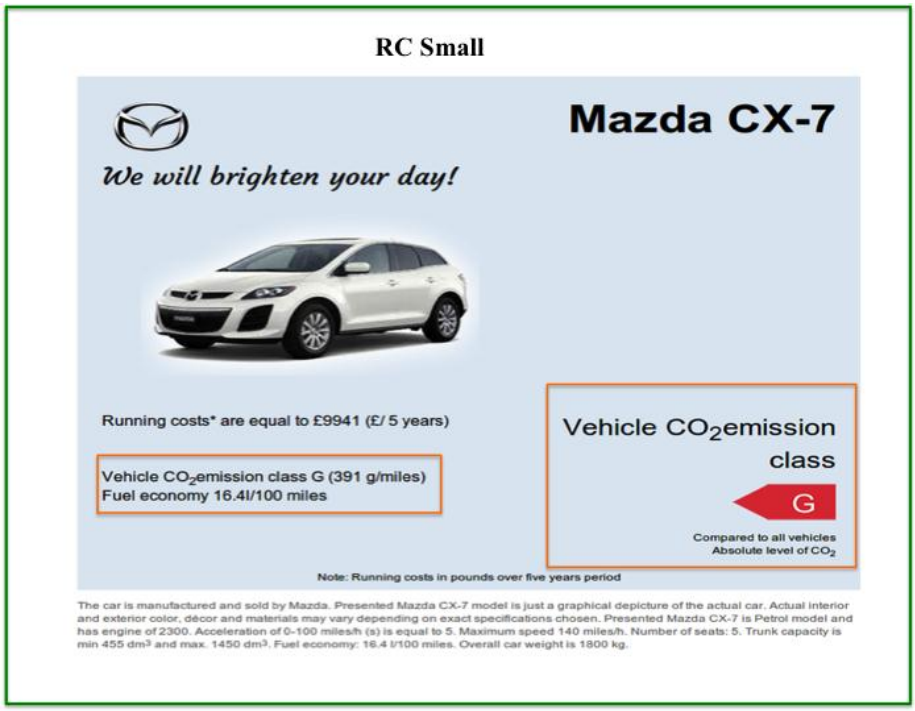
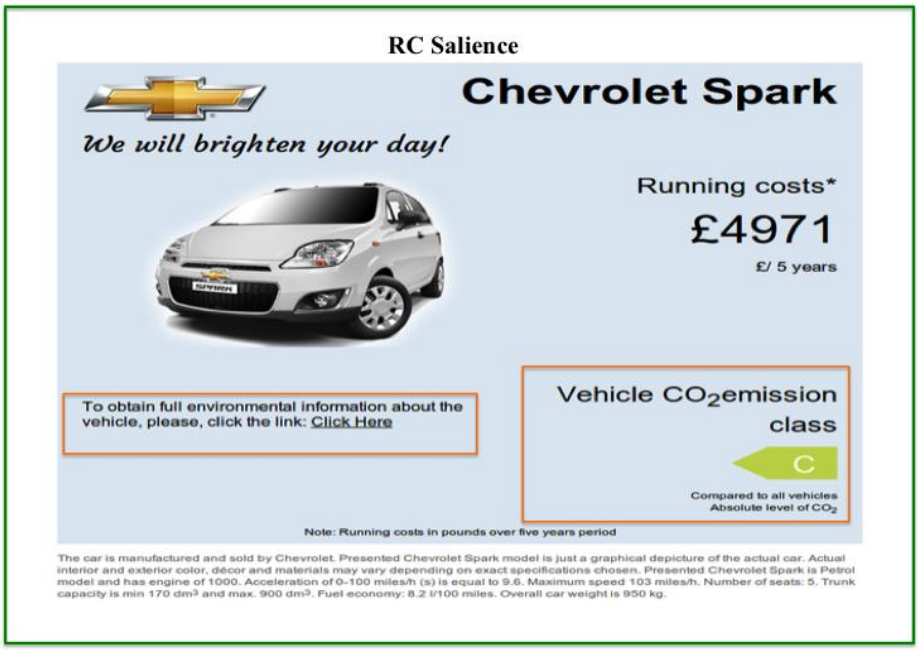
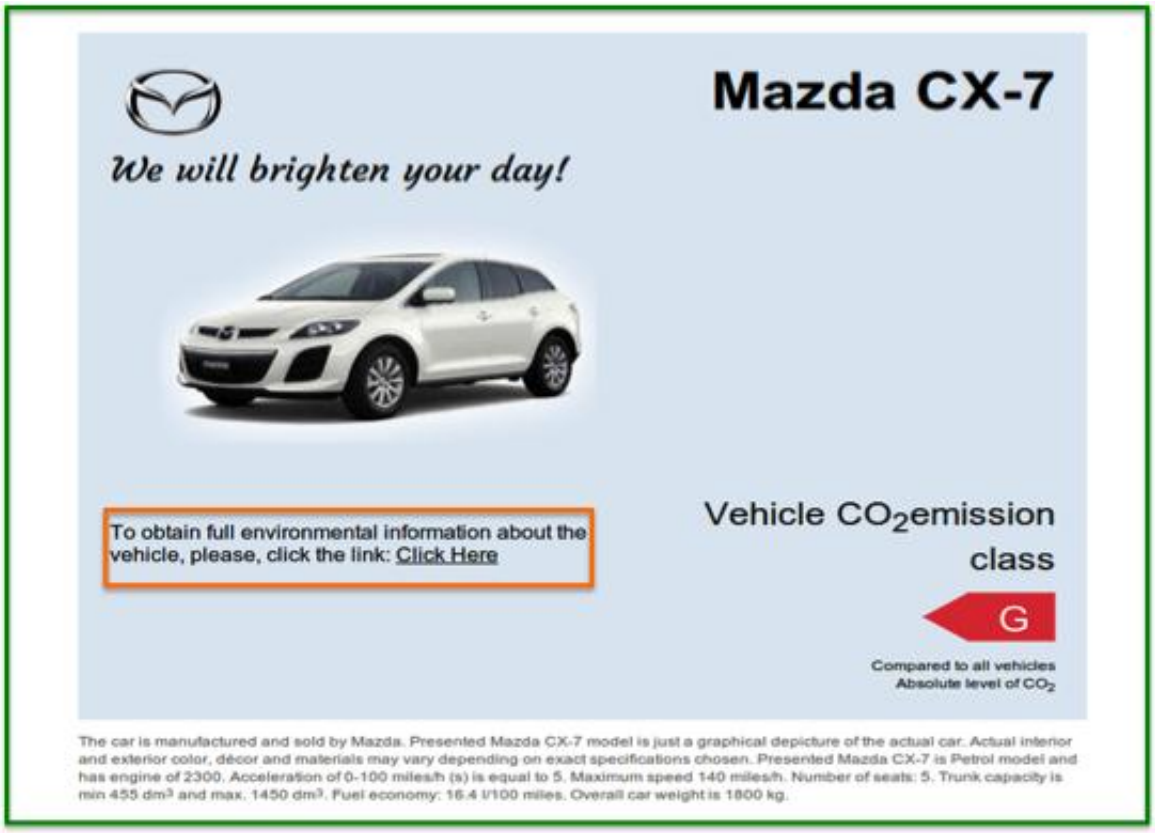


Figure 32 Additional element (promotional material, online experiment) 3



- **Web link**: Present or absent;

Figure 33 Web link (promotional material, online experiment)



Selected references:

Blend, J. R., & Van Ravenswaay, E. O. (1999). Measuring consumer demand for ecolabeled apples. *American Journal of Agricultural Economics, 81*(5), 1072-1077.

Branningan, C., Skinner, I., Gibson, G., & Kay, D. (2011). *Report on the implementation of the Directive 1999/94/EC relating to the availability of consumer information on fuel economy and CO2 emissions in respect of the marketing of new passengers car*. Brussels: Report prepared by ADAE for the European Commission, DG Climate Action.

Brohmann, B., et al. (2009). *What's Driving Sustainable Energy Consumption? A Survey of the Empirical Literature*: Discussion Paper No. 09-013 for the Centre for European Economic Research (ZEW).

Clark, C. F., Kotchen, M. J., & Moore, M. R. (2003). Internal and external influences on pro-environmental behavior: Participation in a green electricity program. *Journal of Environmental Psychology, 23*(3), 237-246.

Department for Transport. (2004). *Assessing the impact of graduated vehicle excise duty: quantitative report*. London: Department for Transport.

Dunlap, R., Van Liere, K., Mertig, A., & Jones, R. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of social issues, 56*(3), 425-442.

Grunig, M., Skinner, I., Kong, M., & Boteler, B. (2010). *Study on consumer information of fuel economy and CO2 emissions of new passenger cars.* Brussels: Report prepared by Ecologic, BIO and IEEP for the European Parliament's Committee on the Environment, Public Health and Food Safety (ENVI).

Johnston, R. J., Wessells, C. R., Donath, H., & Asche, F. (2001). Measuring Consumer Preferences for Ecolabeled Seafood: An International Comparison. *Journal of Agricultural and Resource Economics, 26*(1), 20-39.

Kahneman, D. (2011). *Thinking fast and slow*. London: Penguin Books.

Lane, B., Banks, N., & Anable, J. (2012). *LowCVP Car Buyer Survey: Testing alternative fuel economy labels*. London: Low Carbon Vehicle Partnership.

Loureiro, M. L., McCluskey, J. J., & Mittelhammer, R. C. (2001). Assessing Consumer Preferences for Organic, Eco-labeled, and Regular Apples. *Journal of Agricultural and Resource Economics, 26*(2), 404-416.

Moon, W., Florkowski, W. J., Br√ºckner, B., & Schonhof, I. (2002). Willingness to pay for environmental practices: Implications for eco-labeling. *Land Economics, 78*(1), 88-102.

Rice, G. (2006). Pro-environmental behavior in Egypt: Is there a role for Islamic environmental ethics? *Journal of Business Ethics, 65*(4), 373-390.

Samson, A., & Voyer, B. (2012). Two minds, three ways: dual system and dual process models in consumer psychology. *AMS Review*, 1-24.

Song, H., & Schwarz, N. (2008). If It"s Hard to Read, It"s Hard to Do. *Psychological Science,, 19*(10), 986.

Strack, F., Werth, L., & Deutsch, R. (2006). Reflective and impulsive determinants of consumer behavior. Journal of Consumer Psychology, 16(3), 205–216

Strathman, A., Gleicher, F., Boninger, D., & Edwards, S. (1994). The Consideration of Future Consequences: Weighing Immediate and Distant Outcomes of Behavior. *J Pers Soc Psychol, 66*(4), 742-752.

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.

Thaler, R., & Sunstein, C. (2008). *Nudge : improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Thaler, R. (1985). Mental accounting and consumer choice. Marketing Science, 4, 199–214

Yates, L. (2009). *Green expectations: Consumers' understanding of green claims in advertising*. London: Consumer Focus, Department for Business, Innovation and Skills (BIS).