



WAGENINGEN
UNIVERSITY & RESEARCH

The Role of Familiarity in Front-of-Pack Label Evaluation and Use: A Comparison between the United Kingdom and The Netherlands

Food Quality and Preference

Herpen, E.; Seiss, E.; Trijp, J.C.M.

<https://doi.org/10.1016/j.foodqual.2012.03.003>

This publication is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this publication please contact openscience.library@wur.nl



Contents lists available at SciVerse ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

The role of familiarity in front-of-pack label evaluation and use: A comparison between the United Kingdom and The Netherlands

Erica van Herpen^{a,*}, Ellen Seiss^b, Hans C.M. van Trijp^{c,1}^a Marketing and Consumer Behavior Group, Wageningen University, 6706 KN Wageningen, The Netherlands^b School of Psychology, University of Surrey, The Netherlands^c Wageningen University, 6706 KN Wageningen, The Netherlands

ARTICLE INFO

Article history:

Received 14 November 2011

Received in revised form 12 March 2012

Accepted 12 March 2012

Available online 17 March 2012

Keywords:

Labeling

Logo

GDA

Multiple traffic light

Nutrition label

Food choice

ABSTRACT

Three labeling schemes – signpost logos, multiple traffic light (MTL) labels, and labels communicating guideline daily amounts (GDA) – dominate the debate on front-of-pack nutrition labeling used to assist consumers in making informed food choices. Although the performance of these labeling schemes has been studied extensively, this has mainly been done with a focus on single labeling schemes within single countries where these labels have already a foothold in the market place. Such a priori familiarity raises issues regarding the generalization of results to other contexts and countries. The present study compares consumer evaluation of nutrition labeling schemes, product choices, and inferred product healthfulness across two markets (UK and the Netherlands) with different front-of-pack labeling histories. Results show that familiarity with the labeling scheme affects self-reported evaluations and usage intentions, but that all labeling schemes are equally effective in stimulating healthful choices. The study further shows evidence that all labels increase the perceived healthfulness of more healthful options and that only MTL and GDA reduce healthfulness perceptions of the less healthful options within an assortment. These results are a first step in further elucidating the underlying cognitive processes involved in consumer evaluation and use of front-of-pack nutrition labeling.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Front-of-pack nutrition labeling has taken off in many countries as a means to assist consumers in making more informed, and thereby more healthful, food choices (Storcksdieck genannt Bonsmann, Fernández Celemin, & Grunert, 2010). Yet, despite their wide presence and the sharp increase in studies on consumer responsiveness to front-of-pack nutrition labels in recent years (for reviews see Drichoutis, Lazaridis, & Nayga, 2006; Grunert & Wills, 2007; Mhurchu & Gorton, 2007), there is still no consensus on a uniform global or regional nutrition label. Many studies focus on the effectiveness of a single labeling scheme (e.g., Balcombe, Fraser, & Di Falco, 2010; Sacks, Rayner, & Swinburn, 2009; Vyth et al., 2009), usually of a label that is well-established in the country in which the study took place or one that was being introduced. Specifically, research into the 'Choices Logo' is dominated by studies in the Netherlands, the 'Green Keyhole' by studies in Sweden, and research into traffic light labels by studies in the UK. This is understandable in terms of effective research design, but carries a problem in terms of the generalization of results. After all, it is

impossible to disentangle whether positive consumer evaluations and choice effects are due to familiarity with the labeling scheme or to specific features of the label. As consumer evaluation and use of nutrition labels are affected by both label and consumer characteristics (Bialkova & Van Trijp, 2010; Visschers, Hess, & Siegrist, 2010), familiarity with labeling schemes is likely to matter. In fact, Grunert and Wills (2007) report that research by the European Food Information Council has "led to the impression that (some) consumers seek out labels actively in the UK, but less so in the other countries, which may be related to the differences in availability of and familiarity with labels that exist between these countries."

Only a limited number of studies have compared alternative labeling schemes in their ability to enhance healthful choices. Some of these have been conducted in countries with limited pre-exposure to any front-of-pack nutrition label at the time of data collection (e.g., Germany; Borgmeier & Westenhoefer, 2009), and other studies have familiarized participants with the basic features of the labeling schemes (e.g., Van Herpen & Van Trijp, 2011). Although this cannot fully compensate for the effects of familiarity with labeling schemes, it allows for a more objective assessment of nutrition label effectiveness on the basis of its features and format only.

The present study investigates the role of familiarity by comparing consumer responsiveness to three basic types of labeling schemes – Choices Logo, multiple traffic light (MTL) label, and

* Corresponding author. Tel.: +31 317 483385; fax: +31 317 484361.

E-mail addresses: Erica.vanHerpen@wur.nl (E. van Herpen), E.Seiss@surrey.ac.uk (E. Seiss), Hans.vanTrijp@wur.nl (H.C.M. van Trijp).¹ Tel.: +31 317 483385; fax: +31 317 484361.

monochrome guideline daily amount (GDA) label, see Appendix A – in two countries (UK and the Netherlands). The GDA label has a high presence in both markets, whereas the Choices Logo is prevalent in the Dutch market and the MTL label in the UK market (Storcksdieck genannt Bonsmann, Fernández Celemin, Larrañaga, et al., 2010). Consumers in both countries thus have had extensive exposure to different labeling schemes in naturally occurring choice situations, making the cross-country comparison an ideal situation to examine the effect of long-term familiarity with labeling schemes. To ensure that country differences can indeed be attributed to familiarity, and not to any other coincidental difference between the countries, individual differences in familiarity with different labeling schemes are investigated as (full) mediators. This extends existing research on consumer evaluation and use of front-of-pack labeling schemes in important ways.

The main contribution lies in the investigation, in a mediated moderation analysis, of the extent to which familiarity with a labeling scheme can account for differences observed between the UK and Dutch market in consumer responsiveness to these labeling schemes. If familiarity with labeling schemes indeed (fully) mediates differences in how consumers from the two countries respond to the labeling schemes, this would indicate that past history with the labels is indeed the underlying reason for these effects. In order to investigate this, the current study examines three types of dependent variables: (a) evaluation of the labeling scheme (Do consumers like the labeling scheme?), (b) product attractiveness and choice (Does the labeling scheme affect product choices?), and (c) inferred product healthfulness (Does the labeling scheme affect how consumers evaluate the healthfulness of products?). The investigation of country differences can help to determine to what extent familiarity with a labeling scheme affects these diverse variables. Furthermore, in addition to our main objective, the present study also investigates the effect of labeling schemes on attractiveness and healthfulness perceptions in more detail. We examine whether or not particular labeling schemes not only favor the more healthful options in the assortment but also “penalize” the less healthful options in the consumer’s perception.

2. Different front-of-pack labeling schemes

2.1. Multi-label comparisons and consumer responsiveness

Different types of labeling schemes, specifically signpost logos, multiple traffic light (MTL) labels, and labels communicating guideline daily amounts (GDA), lead the debate on front-of-pack nutrition labeling. The labeling schemes represent three basic types of labels, with varying degrees of directiveness (Hodgkins et al., 2011; Storcksdieck genannt Bonsmann, Fernández Celemin, & Grunert, 2010). Signpost logos provide an overall “approval mark” to healthful products; MTL labels use colors to communicate whether specific nutrients are present in low, medium, or high quantities; and GDA labels provide numerical details per nutrient. Three key issues seem to dominate the debate. The first is whether nutrition labels should communicate product healthfulness in an all or nothing format, as signpost logos do, or also in terms of the underlying nutrient content profiles, as MTL and GDA labels do. A second issue is whether the labeling scheme should provide information on all foods in an assortment, or should only appear on healthful options. And finally, a third issue is whether color coding helps consumers to interpret the nutrient content of products, or whether such information should be provided in more analytical terms (in % daily allowance). These issues are clearly interrelated as signpost logos selectively “favor” the healthful options, but not explicitly disqualify less healthful options. MTL labels, in con-

trast, signal both the healthful options (in terms of green colors) and the less healthful options (in terms of red colors). GDA labels are similarly symmetric but in a more indirect numerical format.

Table 1 provides an overview of published scientific studies that have empirically compared different front-of-pack labeling schemes for packaged food products. It shows that the MTL and GDA labels have been most often compared in various studies. Many of these studies appear to favor color coding of nutrient profiles, albeit that most were conducted in countries outside Europe. Consumers tend to prefer color coded labeling schemes over monochrome labels (Gorton, Mhurchu, Chen, & Dixon, 2009; Kelly et al., 2009; Maubach & Hoek, 2010) and the use of color coding increases the correct identification of more as well as less healthful alternatives (Borgmeier & Westenhoefer, 2009; Kelly et al., 2009; Maubach & Hoek, 2008), although other studies find that both color coded labeling and monochrome labeling are understood well (Grunert, Wills, & Fernández-Celemin, 2010). An increased ability of consumers to differentiate between more and less healthful products may, however, also lead to more extreme product evaluations than is warranted by the underlying product differences (Balcombe et al., 2010; Hieke, Wilczynski, & Schwaiger, 2011). Furthermore, logos can be helpful in increasing brand attitudes (Kozup, Creyer, & Burton, 2003), perceived healthfulness of moderately healthful products (Andrews, Burton, & Kees, 2011), and healthful choices (Van Herpen & Van Trijp, 2011).

2.2. Cross-country comparisons on front-of-pack labeling effectiveness

Only a limited number of studies have been conducted in the cross-country context (see Table 1). Still, although not formally tested, the empirical results of several of the studies included in Table 1 suggest that cross-country differences in familiarity may affect how consumers evaluate and use nutrition labels. For instance, the study of Grunert, Fernández-Celemin, et al. (2010) has uncovered considerable cross-national differences in the use and understanding of nutrition labels. Their study shows that the GDA label is used more often in the UK, France, and Germany than in the other countries (Sweden, Poland and Hungary), whereas consumers only reported using color-coded systems in the UK and France – the two countries in which these systems were present at the time of the study. They also found country differences in the nutrients that consumers look for on product packages. This appears to reflect the attention given to specific nutrients in media outlets, as consumers in the UK, where recent national campaigns about salt and saturated fat took place, paid relatively more attention to these nutrients than consumers from other countries.

Additionally, Möser, Hoefkens, Van Camp, and Verbeke (2010) compared the MTL and GDA labels by directly asking consumers to evaluate these labeling schemes. Their results show that German consumers think that reading a MTL label is less time-consuming than reading a GDA label, whereas Belgian consumers (although on another scale) interpret the GDA label much more positively than the MTL label. The authors argue that these differences in label perception may possibly result from the ongoing public discussion on the introduction of the MTL label in Germany. This public discussion has brought the MTL label to the attention of German consumers, whereas in Belgium there is far less public attention for the MTL label. There, the GDA label has been widely introduced and generally accepted. Country differences have also been reported in the study by Feunekes, Gortemaker, Willems, Lion, and Van den Kommer (2008), although the pattern there is less clear-cut and not interpreted further by the authors.

In conclusion, although previous studies have identified country differences in consumer responsiveness to front-of-pack nutrition labels, to the best of our knowledge no prior study has formally explored the mediating role of label familiarity in this process. We

Table 1
Studies comparing different front-of-pack labeling schemes.

Paper	Labeling schemes ¹	Dependent variables	Main results	Country
Andrews, Burton, and Kees (2011)	Choices logo, CC-GDA	Product healthfulness, nutrient evaluation, attitude, purchase intention; for moderately healthful products	Logo increases perceived healthfulness and decreases perceived levels for 'negative' nutrients more than CC-GDA. Both labels increase product attitude and purchase intentions	US
Bialkova and Van Trijp (2010)	Choices Logo, MC-GDA, CC-GDA	Detection of label presence on product	No significant differences for single label detection. Attention capture faster for choices logo than GDA when asked whether one or two labels are present	The Netherlands
Bialkova and Van Trijp (2011)	MC-GDA, CC-GDA	Product choice, response time	MC-GDA increases choice likelihood compared to CC-GDA, for preference goal	The Netherlands
Borgmeier & Westenhoefer 2009	Tick, MTL, MC-GDA, CC-GDA	Identification of healthful alternative	Correct identification highest for MTL, followed by MC-GDA and CC-GDA	Germany
Feunekes et al. (2008), study 1	Tick, Health protection factor, Stars, Smileys, MTL, WTL	Evaluation and attractiveness of labeling scheme, product healthfulness	In general, MTL and WTL evaluated best. Smileys and Stars are best discriminators between healthful and less healthful products	UK, Germany, Italy, the Netherlands
Feunekes et al. (2008), study 2	Tick, Stars, Multiple choice tick, MC-GDA	Evaluation and attractiveness of labels, product healthfulness, intended product usage frequency, evaluation time	GDA and Stars liked most. No difference for intended usage frequency. GDA requires more time to evaluate	Italy, UK
Gorton et al. (2009)	MTL, STL, MC-GDA, NIP	Product healthfulness, attractiveness of labeling scheme	MTL preferred most, GDA preferred least. Product healthfulness most accurate with STL, followed by MTL; with NIP and GDA scoring worst	New Zealand
Grunert, Fernández-Celemin, et al. (2010)	MC-GDA, CC-GDA, MTL, NIP, Logo	Looking at labels, reported looking at labels, understanding of GDA label	Use of nutrition labels varies per country and product category. MTL and CC-GDA used most often in UK, MC-GDA in UK, France and Germany, health logo in Poland and Hungary. Application of GDA label in relative product judgments is good in all countries	UK, Sweden, France, Germany, Poland, Hungary
Grunert, Wills, and Fernández-Celemin (2010), study 1	GDA, MTL, NIP	Looking at the label, reported looking at labels	Mentioning of specific labels depends on retailer adoption of these labels	UK
Grunert, Wills, and Fernández-Celemin (2010), study 2	MC-GDA, MTL, CC-GDA	Understanding of the label	Subjective understanding not significantly different. Understanding high for all labels	UK
Jones & Richardson 2007	NIP, NIP plus MTL	Product healthfulness, attention to the label and nutrients	MTL affects which nutrients are attended to and leads to more correct healthfulness ratings	UK
Kelly et al. (2009)	MC-GDA, CC-GDA, MTL, MTL plus overall rating	Attractiveness of labeling scheme, identification of healthful product	CC-GDA preferred most, followed by MTL plus rating and MTL. Product identification most accurate for MTL and MTL plus rating; less accurate for CC-GDA and GDA	Australia
Kozup et al. (2003)	NIP, NIP plus heart healthy symbol and claim	Brand attitude, purchase intention	The heart healthy symbol and claim increases brand attitudes	US
Maubach (2008)	MC-GDA, MTL, NIP	Product attitude	MTL better able to enhance accurate evaluation of products than GDA. MTL helps identify less healthful products	New Zealand
Maubach, Hoek, Healey, Gendall, and Hedderley (2009)	MTL, MC-GDA	Product choice	MTL enhances discrimination between healthful and less healthful products more than GDA	New Zealand
Maubach and Hoek (2010)	MC-GDA, MTL, WTL, CC-GDA, NIP	Evaluation and attractiveness of the labeling scheme	Evaluation of color coded labels more positive than of wholly numeric labels	New Zealand
Möser et al. (2010)	MC-GDA, MTL	Attractiveness of labeling scheme	In Belgium, GDA preferred. In Germany, MTL preferred	Germany and Belgium
Scott and Worseley (1994)	NIP, Tick, Healthy Food Pyramid, Hearth foundation claim	Appeal of labeling scheme, understanding of labels	NIP was least appealing; Tick and claim were most appealing. Some level of misunderstanding reported for all labels	New Zealand
Van Herpen and Van Trijp (2011), study 1	Tick, MTL, NIP	Appeal of labeling scheme, attention to label, label recognition, product choice	NIP has higher appeal than MTL and Tick. MTL and especially Tick enhance choice of healthful products	The Netherlands
Van Herpen and Van Trijp (2011), study 2	Tick, MTL, NIP	Appeal of labeling scheme, attention to label, label recognition, product choice	NIP has higher appeal than MTL and Tick. MTL and Tick enhance choice of healthful products, even under time pressure	Turkey

¹ MTL, multiple traffic light label; MC-GDA, monochrome label based on guideline daily amounts in some studies referred to as % DI (% daily intake); CC-GDA, color coded label based on guideline daily amounts; NIP, (shortened) nutrition information panel on front-of-pack; WTL, wheel traffic light; STL, simple traffic light (one overall traffic light rating for the product).

build on previous studies to develop specific hypotheses on how labels may be evaluated and used differently in different countries, before we formally test the mediating role of familiarity as an explanation of country differences between the UK and the Netherlands.

2.3. Evaluation of the labeling schemes

Consumers often have strong ideas on which labels they like or do not like. Yet, these evaluations of labeling schemes are often not in line with actual label use (Levy, Fein, & Schucker, 1992), primar-

ily because consumers tend to underestimate the effort required in using labels with more detailed information (Drichoutis et al., 2006; Grunert & Wills, 2007). Thus, although consumers acknowledge that nutrition tables are relatively more difficult to use, they tend to evaluate these labels more positively than MTL labels and signpost logos (Van Herpen & Van Trijp, 2011). We expect that this finding extends from nutrition tables printed front-of-pack to monochrome GDAs even when they provide some direction. Hence we hypothesize:

H1: The monochrome GDA label is evaluated more positively than the MTL label and Choices Logo, with the exception of difficulty in use.

Additionally, based on the review of the literature, we assume that the “home” label that consumers are familiar with will be evaluated more positively than the unknown label. In other words, consumers are likely to evaluate labels better when they already have some experience with them. In the present context of a comparison between the UK and the Netherlands, the monochrome GDA label is present in both countries, whereas the MTL is relatively more prominent in the UK and the Choices Logo is relatively more prominent in the Netherlands. We expect that:

H2: In the UK, the MTL label is evaluated more positively than the Choices Logo, whereas in the Netherlands the opposite is true.

2.4. Healthful choices

Even though consumer evaluations of the different labeling schemes may be a poor predictor of their effectiveness in guiding healthful choices, nutritional labeling schemes provide relevant guidance in facilitating the healthful choice, simply because of the transparency that they provide. Hence, we hypothesize that:

H3: Nutrition labels front-of-pack are effective in stimulating the more healthful choice, compared to situations where front-of-pack nutrition labels are absent.

In addition, we expect that the degree to which front-of-pack nutrition labels enhance healthful choices depends on the prominence with which these labeling schemes appear in daily life. After all, consumers may become more competent in using a labeling scheme when they are more familiar with it. Hence:

H4: The Choices Logo (MTL label) is more effective in stimulating the more healthful choice in the Netherlands (UK), and the monochrome GDA label is equally effective in the Netherlands and the UK.

Finally, we formally test whether familiarity with the labeling scheme mediates the effectiveness of alternative labeling schemes in stimulating the healthful choice.

H5: The effect of alternative labeling schemes on (a) the evaluation of labeling schemes and (b) the likelihood of making the more healthful choice is mediated by familiarity with the labeling scheme.

2.5. How labels enhance healthful choices

Different labeling schemes take a different approach to stimulating healthful food choices and this has been central in the debate among various stakeholders (Bussell, 2005). Whereas signpost logos are awarded only to the more healthful food prod-

ucts within the assortment, color-coded labels (MTL label and color-coded GDA not only signify the relatively healthful options (predominantly green/amber colored labels) but also identify the less healthful options (predominantly red/amber colored labels). This contributes to increasing transparency and differentiation within the assortment. Monochrome GDA labels achieve the same objective but are less salient within the choice assortment. Different stakeholders have advocated different labeling schemes, with the health sector mainly favoring the MTL label and food manufacturers the GDA label (White, Thomson, & Signal, 2010), and all labeling schemes have been both advocated and highly criticized (see e.g., Bussell, 2005; Halliday, 2009).

Because all labeling schemes are designed to signify the healthful options in the assortment, we expect all labeling schemes to strengthen the healthfulness perception of the more healthful options within the assortment, hence:

H6: All labeling schemes increase (a) the perceived healthfulness and (b) the attractiveness of more healthful products, compared to the no-label condition.

Additionally, based on the label design, we expect MTL and GDA labels to also decrease perceived healthfulness and attractiveness of the less healthful options. This effect may be especially large for the MTL label, as consumers have a tendency to overinterpret the meaning of amber and red colors in this label (Grunert, Wills, & Fernández-Celemin, 2010). For instance, they tend to interpret a red color as a signal that they should try to not eat the product, whereas the correct interpretation is that it is fine to have such a product occasionally as a treat. Consumers consequently demonstrate a strong aversion to food products displaying red traffic lights (Balcombe et al., 2010; Hieke et al., 2011). In contrast, we expect that asymmetric labeling schemes (e.g., Choices Logo) leave the perceived healthfulness of the less healthful options relatively unaffected. Hence:

H7: MTL and GDA labels decrease (a) the perceived healthfulness and (b) the attractiveness of less healthful products, whereas the Choices Logo does not.

As prior research has shown that shopping goals are very important as a top-down factor driving consumer response to and evaluation of front-of-pack nutrition labels (e.g., Van Herpen & Van Trijp, 2011; Visschers et al., 2010), we also incorporate specific shopping goals in the research design without developing formal hypotheses. We expect replication of the results found in these prior studies.

3. Experiment: Nutrition labels and familiarity

3.1. Method

3.1.1. Participants and design

Participants were 186 students and administrative staff members below 30 years old from a UK University (69% female, mean age 22.8 years) and 197 students and administrative staff members from a Dutch University (63% female, mean age 20.3 years). People with relevant food allergies or color blindness were excluded. After giving informed consent, participants were randomly assigned to one of the conditions in a 4 (labeling scheme: none, Choices Logo, MTL, monochrome GDA (MC-GDA)) \times 3 (goal: preference, general health, low saturated fat) between subjects design. Between countries, the samples did not differ in gender distribution. The UK sample was slightly, albeit significantly, older than the Dutch sample (23 vs. 20 years, $F(1, 382) = 80.00, p < .001$).

3.1.2. Labeling schemes

Participants were exposed to one of four labeling schemes. The ‘no label’ condition presented the products without any nutrition labeling. The other labels (Choices Logo, MTL, MC-GDA) were presented in ecologically valid size. In line with existing market condition, the Choices Logo was only present on the two pizzas that would qualify against the Choices criteria, whereas the other two labeling schemes appeared on all options. Further, to increase realism, both the MTL and GDA (but not the Choices Logo) were complemented with a calorie label front-of-pack, as indicated in Appendix A. In reality, calories are commonly provided as part of, or in addition to, the nutrient information contained in GDA and MTL labels. Calorie labeling front-of-pack can be an important cue for inferences about product healthfulness (Van Kleef, Van Trijp, Paeps, & Fernández-Celemin, 2008).

3.1.3. Goal manipulation

The task instructions were (preference condition): “Imagine the following situation. You have planned to go to the supermarket to buy pizza. On the next screen you will see different pizzas. Please look carefully at all the options. Which pizza would you buy? When you have made your choice, click on the chosen pizza to continue.” Alternatively (health and low saturated fat conditions): “Imagine the following situation. You have decided to eat healthier/eat less saturated fat. You have planned to go to the supermarket to buy pizza. On the next screen you will see different pizzas. Please look carefully at all the options. Which pizza would you buy? Keep in mind that you have decided to eat healthier/eat less saturated fat. When you have made your choice, click on the chosen pizza to continue.”

3.1.4. Stimuli

The product category of pizza was used for this study, because healthfulness can differ substantially between options from this category, it is a main meal, and it is a product category that participants are familiar with. Six types of pizza were selected: Venezia, Toscana, Bolzano, Portofino, Caprese, and Romagna. These pizza types exist in practice, but do not immediately detail the ingredients, preventing participants from using healthfulness information stored in memory. For the same reason, a fictitious brand name was used. The pizzas were in reality pizzas from a Dutch retail brand, but with references to the store removed and a new brand name (Delizioso) included as well as nutrition labels (depending on condition), as shown in Appendix A.

The pizzas were experimentally designed to represent variations in nutrient profile and hence partial (GDA and MTL) and overall (Logo) level of healthfulness. Existing nutrient profiles of several commercial pizzas were examined to ensure that profiles were representative of products in the market. Based on the nutrient content (see Table 2), choices were classified as more healthful (“Portofino” or “Romagna”), medium healthful (“Toscana” or “Bolzano”), or less healthful (“Venezia” or “Caprese”). The more healthful options outperformed the other pizzas on all indicated nutrients and energy, and the medium healthful options outperformed the less healthful options on all indicated nutrients and energy.

3.1.5. Procedure

Data was collected on computers using the software package Authorware. The two countries in which data were collected differ in terms of presence (and thus familiarity) with the nutrition labels. Hence, participants received information on screen which briefly explained the labeling schemes to bring their understanding to an adequate level. This was presented in the form of a leaflet (see Appendix B). To enhance task realism, participants evaluated

the quality of the leaflet, as well as the three labels described in the leaflet.

In a choice task participants subsequently saw on screen a picture of an assortment with six pizzas. A large screen (27 inch) was used to ensure that all information on the product packages was easily readable. Pizza choice was recorded by the computer. After choice participants were asked an open-ended question on why they selected this pizza. Next, they received instructions that they would rate each of the pizzas in terms of attractiveness and overall healthfulness. To refresh their memory and allow participants to correctly differentiate between the products (which product was which), the assortment of pizzas was shown on screen for 10 s, after which the ratings were asked. The assortment was not visible on screen when the ratings were provided, to prevent participants from focusing on information that they would not naturally notice. Finally, participants filled in a questionnaire collecting background characteristics, were debriefed and thanked. Participation lasted around 15–20 min and participants received a free product (chocolate bar, bag of nuts, etc.) as a reward.

3.1.6. Measures

Quality of the leaflet was measured with two items, rated on 7-point scales: “How clear did you find the leaflet?” (not at all clear – very clear) and “How interesting did you find the leaflet?” (not at all interesting – very interesting).

Evaluation of the labeling schemes was collected on the basis of the leaflet, using five items for each of the three labeling schemes. These were rated on 7-point scales: “How clear do you find the labels in the leaflet?” (not at all clear – very clear), “How difficult would it be to use the labels when choosing food products?” (not at all difficult – very difficult), “How much do you trust each of the labels?” (not at all – very much), “To what extent do you find the labels patronizing?” (not at all patronizing – very patronizing), and “Would you use these labels when making a product choice” (definitely would not – definitely would).

Familiarity with the labeling schemes was measured directly following the exposure to the leaflet (cf. Van Herpen & Van Trijp 2011). Participants saw a picture of each of the labels in random order and indicated if they had seen this type of label before it was described in the leaflet, on a three point scale (no, I have never seen it before/no, but I have seen something like it/yes, I have seen this label before). The middle category was used to ensure that familiarity with a slightly different version of a specific type of label (e.g., a vertically oriented MTL label) was also taken into account. In the analyses, the latter two categories were combined to obtain a dichotomous measure of familiarity with the label.

Healthful choice was derived from the chosen pizza, as indicated in Section 3.1.4.

Choice motivation was measured as an open-ended question immediately after the choice task: “You have made a choice between different types of pizzas. On what did you base your choice?” Answers were content analyzed, and presence or absence of label as a reason for making the choice was coded. Open responses were only coded as “label” when there was explicit mentioning of the label or label elements (e.g., green light, percentage, etc.), not if only a nutrient (e.g., fat) was mentioned.

Perceived healthfulness was asked for each pizza, on 7-point semantic differential scales with endpoints “very unhealthy” and “very healthy”. Pizzas were indicated by their name and position in the assortment (e.g., Venezia (left, up)), and were not visible on screen while ratings were provided. Healthfulness ratings for the two more healthful pizzas were averaged, as were those for the two medium healthful and for the two less healthful pizzas.

Product attractiveness was also asked for each of the pizzas, on 7-point semantic differential scales with endpoints “I definitely do not prefer this pizza” to “I definitely prefer this pizza”. Pizzas

Table 2
Nutrient profiles for the pizzas.

Pizza	Ingredients	Nutrient	Grams per 100 g	% GDA	MTL color	Calories per 100 g	Logo
Venezia	Mozzarella, squid, shrimps, smoked salmon, clams, mussels and olives	Sugar	12.3	14	Amber	432	No
		Fat	20.0	29	Amber		
		Sat. fat	7.2	36	Red		
		Salt	2.3	96	Red		
Toscana	Mozzarella, spicy gorgonzola, roasted paprika, eggplant, zucchini, green asparagus and broccoli	Sugar	8.2	9	Amber	321	No
		Fat	11.0	16	Amber		
		Sat. fat	3.2	16	Amber		
		Salt	0.70	29	Amber		
Bolzano	Mozzarella, spicy gorgonzola and goat cheese, feta and roasted paprika	Sugar	8.8	10	Amber	344	No
		Fat	13.2	19	Amber		
		Sat. fat	3.6	18	Amber		
		Salt	0.82	34	Amber		
Portofino	Spinach, tomato and pine nuts	Sugar	5.1	6	Amber	239	Yes
		Fat	3.0	4	Green		
		Sat. fat	1.5	8	Green		
		Salt	0.35	15	Amber		
Caprese	Mozzarella, buffalo mozzarella and cherry tomatoes	Sugar	11.8	13	Amber	420	No
		Fat	21.2	30	Red		
		Sat. fat	6.2	31	Red		
		Salt	1.5	63	Amber		
Romagna	Roasted chicken, mozzarella, onion and paprika	Sugar	2.7	3	Green	203	Yes
		Fat	3.4	5	Amber		
		Sat. fat	1.4	7	Green		
		Salt	0.31	13	Amber		

were again indicated by name and position and were not visible on screen while ratings were provided. Attractiveness ratings were averaged for the more healthful, for the medium healthful, and for the less healthful pizzas.

Background characteristics included socio-demographics (age, gender, and education program engaged in), use of pizzas in the dietary pattern (“How often do you eat pizza” on a 7-point scale ranging from “never” to “always”), and liking of pizza (“Do you like the taste of pizza” on a 7-point scale ranging from “do not like at all” to “like very much”). These were taken to ensure that the random assignment of participants was successful, and we indeed found that these variables did not significantly differ between conditions.

3.1.7. Data analysis

Differences between the countries in terms of background variables and familiarity with labeling schemes were tested using ANOVA's and χ^2 tests, as appropriate. As the same participant evaluated all three labeling schemes, repeated measures ANOVA was applied for labeling scheme evaluations, with as independent variables: label (as a within-subjects factor), country (as a between-subjects factor), and the interaction between label and country. Label differences were investigated using contrast tests with dummy (−1,0,1) coding. If significant interactions were found, mediated moderation was tested for, using familiarity with the label as the mediator and following the procedure outlined by Muller, Judd, and Yzerbyt (2005).

Healthful choice was examined using a multinomial logistic regression, with country, task, and labeling scheme as independent variables. Potential interaction effects were investigated, and mediated moderation models would be used to investigate the mediating role of familiarity if country differences were found. Initially, the model used preference task (for task) and the control condition without labels (for labeling scheme) as reference categories, but in follow-up analyses different reference categories were selected, to obtain estimates for differences between health task and low saturated fat task, and between the labeling schemes.

Choice motivation provided in the open answers were content analyzed to identify whether or not the label was mentioned as a motivation (yes/no) for pizza choice and served as dependent variables in a logistic regression. Independent variables were country, task, labeling scheme, together with all 2-way and 3-way interactions. As reference categories, preference task (for task) and GDA label (for labeling scheme) were used, and the control condition was excluded from this analysis.

Healthfulness and attractiveness ratings were provided for all pizzas by any single participant and hence analyzed with repeated measures ANOVA's. Pizza type (more healthful, medium healthful, and less healthful) was the repeated measures factor and labeling scheme, task, and country were between subject factors in these analyses. Differences between labeling schemes were investigated using post hoc (LSD) tests.

3.2. Results

3.2.1. Background

Participants on average liked pizza ($M = 5.66$) and were inclined to eat it on occasion ($M = 3.05$). There were no differences between the countries, neither for liking of pizza ($F(1,381) = 1.21$, *ns*), nor for frequency of eating pizza ($F(1,281) = 0.67$, *ns*). The leaflet introducing the labels was seen as clear ($M = 5.75$) and this holds for both countries ($F(1,381) = 0.12$, *ns*). However, it was rated as marginally more interesting ($F(1,381) = 3.47$, $p = .06$) in the Netherlands ($M = 4.66$) than in the UK ($M = 4.44$).

3.2.2. Familiarity with the labeling schemes

As anticipated, participants in the Netherlands had more experience with the Choices Logo (97.5% had seen this label before) than participants in the UK (28.5% had seen it before) ($\chi^2(2) = 197.45$, $p < .001$). In contrast, participants in the UK had more experience with the MTL label (68.8% had seen this label before) than participants in the Netherlands (22.8% had seen it before) ($\chi^2(2) = 81.65$, $p < .001$). Most participants in both countries were familiar with the GDA label (94.1% and 91.9% in UK and the

Netherlands, respectively) and to the same degree ($\chi^2(2) = 3.45$, *ns*).

3.2.3. Evaluation of the labeling schemes

Table 3 shows the results of the repeated measures ANOVAs. Significant main effects of labeling scheme were present for all variables. Specifically, the GDA label was seen as more difficult ($M = 3.18$) than both the Choices Logo ($M = 2.85$) and MTL label ($M = 2.87$), which did not significantly differ on perceived difficulty. For all other measures, the results confirmed those of Van Herpen and Van Trijp (2011), namely that the Choices Logo was seen as less clear, less trustworthy, more patronizing, and less likely to be used than the non-directive label (GDA in this case), with the MTL label in between. These results were in support of Hypothesis 1.

In addition, there were significant main effects of country on all variables except perceived difficulty to use the label. Participants from the UK found labels more clear ($M_{UK} = 5.18$ vs. $M_{NL} = 4.92$), trusted labels more ($M_{UK} = 4.86$ vs. $M_{NL} = 4.69$), thought labels were less patronizing ($M_{UK} = 3.21$ vs. $M_{NL} = 3.73$), and intended to use them more ($M_{UK} = 4.87$ vs. $M_{NL} = 4.24$).

These results were qualified by significant country \times label interactions. To examine these in more detail, Table 3 indicates results of contrast tests to examine differences between labels for the separate countries. While the patterns for trust and paternalism were similar in both countries and in line with the main effects, there were noticeable differences between the two countries for the other variables. For clearness, participants from the Netherlands evaluated the MTL label as less clear than the GDA label, while participants from the UK showed no significant differences between these two labels. For difficulty, UK participants found the MTL labeling scheme easier to interpret than the Choices Logo. In contrast, Dutch participants found the Choices Logo easiest to interpret. For usage intentions, although in both countries participants claimed highest usage intentions for the GDA label, UK participants claimed lower usage intentions for the Choices Logo compared to the MTL label, whereas in the Netherlands usage intentions for these two labeling schemes were similar. These results are in line with the expectation that consumers from the UK evaluate the MTL label better than the Choices Logo, thus providing partial support for Hypothesis 2.

3.2.4. The mediating role of familiarity in label evaluation

Next, we investigated to what extent the label \times country effects in labeling scheme evaluations were mediated by familiarity with the label. This is an example of mediated moderation. Following

the procedure suggested by Muller et al. (2005), mediated moderation in our case should manifest itself in (a) a significant country \times label interaction on label evaluation, (b) a significant effect of country on familiarity, and (c) a significant interaction between labeling scheme and familiarity on label evaluation; when including this latter interaction, the country \times labeling scheme interaction should become insignificant (“full” mediated moderation) or decrease in size (“partial” mediated moderation). Condition (a) was met for all variables except how clear the label was (where the country \times label interaction is marginally significant) (see Table 3). In line with condition (b), country significantly affected both familiarity with the MTL label and familiarity with the Choices Logo, as reported in the section on familiarity with the labeling schemes.

To examine condition (c), a model was tested in which evaluation of the labeling scheme was explained by label, country, Logo familiarity, MTL familiarity, the country \times label interaction, the Logo familiarity \times label interaction, and the MTL familiarity \times label interaction. Full mediated moderation would be established if the Logo familiarity \times label interaction and/or the MTL familiarity \times label interaction were significant, whereas the country \times label interaction became insignificant. This was indeed the case for two important variables: how difficult it would be to use the label and usage intentions, providing partial support of Hypothesis 5a, while for the other label evaluations (clear, trust, patronizing) no mediated moderation is observed. For these two variables, both the Logo familiarity \times label and MTL familiarity \times label interactions were significant (all $F_s > 3$, $p < .05$), and the country \times label interaction became insignificant ($F(2,758) = 0.72$, *ns*, for difficulty and $F(2,758) = 0.73$, *ns*, for usage intentions). Inspection of the means showed that, as expected, participants who were familiar with the Choices Logo rated this logo as less difficult to use ($M = 2.58$) and intended to use it more ($M = 4.22$) than participants who were not familiar with the logo ($M = 3.35$ and $M = 3.78$, respectively). Similarly, participants who were familiar with the MTL label rated this label as less difficult to use ($M = 2.68$) and intended to use it more ($M = 4.99$) than participants who were not familiar with the MTL label ($M = 3.03$ and $M = 4.11$, respectively). Thus, the country differences found for difficulty of use and usage intentions could be completely accounted for by differences in label familiarity.

3.2.5. Healthful choice

A multinomial logistic regression was used to examine the effect of country, task (preference task serving as reference category), and labeling scheme (control condition without labels serving as reference category) on product choice. The inclusion of interaction

Table 3
Evaluation of the labeling schemes.

		Logo	MTL	GDA	Label	Country	Label \times country
Clear	UK	4.24 _a	5.59 _b	5.72 _b	$F(2,762) = 139.84^{***}$	$F(1,381) = 7.20^{**}$	$F(2,762) = 2.86^\dagger$
	NL	3.88 _a	5.14 _b	5.74 _c			
Difficult	UK	3.12 _a	2.72 _b	2.93 _{ab}	$F(2,762) = 7.14^{**}$	$F(1,381) = 0.64$ <i>ns</i>	$F(2,762) = 16.17^{***}$
	NL	2.60 _a	3.02 _b	3.42 _c			
Trust	UK	4.02 _a	4.89 _b	5.67 _c	$F(2,762) = 211.61^{***}$	$F(1,381) = 1.81^\dagger$	$F(2,762) = 4.56^*$
	NL	3.81 _a	4.48 _b	5.79 _c			
Patronizing	UK	3.72 _a	3.36 _b	2.55 _c	$F(2,762) = 131.41^{***}$	$F(1,381) = 17.19^{***}$	$F(2,762) = 3.50^*$
	NL	4.36 _a	4.02 _b	2.80 _c			
Usage intention	UK	4.06 _a	5.06 _b	5.49 _c	$F(2,762) = 58.54^{***}$	$F(1,381) = 24.39^{***}$	$F(2,762) = 17.37^{***}$
	NL	4.06 _a	3.99 _a	4.68 _b			

Letters indicate which means differ using contrasts, for each country separately.

$^\dagger p < .10$.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

terms was explored, but none reached significance, and the model with main effects only is reported. Likelihood ratio tests compared this latter model with a reduced model in which one of the factors was omitted to investigate whether the factors had a significant contribution. This was indeed the case for both labeling scheme ($\chi^2(6) = 26.42, p < .001$) and task condition ($\chi^2(4) = 17.24, p < .01$), but not for country ($\chi^2(2) = 0.34, ns$). Parameter estimates further confirmed that country did not affect the likelihood of choosing a less healthful product compared to a more healthful product ($\chi^2(1) = 0.85, ns$) nor the likelihood of choosing a medium healthful product compared to a more healthful product ($\chi^2(1) = 0.25, ns$). As including a label \times country interaction also did not produce significant effects ($p = .6$), this implied that although participants had claimed they would use specific labels more depending on the country they resided in, this was not revealed in their product choices. Hence, Hypothesis 4 is partially supported, for GDA but not for the Choices Logo nor for MTL. As a consequence, Hypothesis 5b is not supported as there is no country effect to be potentially mediated by familiarity.

The effects of labeling scheme revealed that the likelihood of choosing a less healthful option compared to a more healthful option decreased when any of the labels was present in comparison to the no label condition (Choices Logo: $\beta = -0.69; \chi^2(1) = 3.58, p = .06$; MTL label: $\beta = -1.79; \chi^2(1) = 15.47, p < .001$; GDA table: $\beta = -1.08; \chi^2(1) = 7.65, p < .01$). Likewise, the likelihood of choosing a medium healthful option compared to a more healthful option decreased when any of the labels was present in comparison to the no label condition (Choices Logo: $\beta = -0.84; \chi^2(1) = 4.35, p < .05$; MTL label: $\beta = -1.41; \chi^2(1) = 10.14, p < .01$; GDA table: $\beta = -1.10; \chi^2(1) = 6.81, p < .01$). In other words, all labeling schemes increased the choice of the more healthful option compared to the control condition without labels, thus providing support for Hypothesis 3.

Analyses using different reference categories revealed that the Choices Logo decreased choice of the less healthful option compared to the more healthful option less than the MTL label did ($\beta = -1.09; \chi^2(1) = 5.62, p < .05$). This may be reflective of a superior ability of the MTL label to signal less healthful options, as we will come back to later. All other differences between labeling schemes were insignificant. Choice percentages are provided in Table 4.

For task, participants were more likely to choose a less healthful product than a more healthful product when given a preference goal, both in comparison to a health goal ($\beta = 1.17; \chi^2(1) = 10.90, p < .01$) and to a low saturated fat goal ($\beta = 1.01; \chi^2(1) = 8.29, p < .01$). Likewise, they were more likely to choose a medium healthful product than a more healthful product when given a preference goal, but only in comparison to a low saturated fat goal ($\beta = 0.87; \chi^2(1) = 4.53, p < .05$). The health goal and low saturated fat goal did not significantly differ in their effectiveness of directing choices. These effects are in line with prior research (Van Herpen & Van Trijp, 2011).

3.2.6. Choice motivation

The extent to which label was mentioned (yes/no) as a motivation for pizza choice (percentages in Table 4) was analyzed using a logistic regression with country, task (preference task as reference category), labeling scheme (GDA label as reference category), and all 2-way and 3-way interactions as independent variables. The control condition without labels was not included (none of participants in this condition mentioned a label). Results showed main effects of labeling scheme ($\chi^2(2) = 12.15, p < .01$) and task condition ($\chi^2(2) = 16.21, p < .001$). In line with the results for healthful choice, neither the effect of country nor any of the interaction effects were significant. Thus, again, we found no evidence that

participants from one country used any of the labeling schemes more or less often than participants of the other country.

Inspection of the effect of labeling scheme showed that spontaneous reference to the label was higher for the MTL label than for the GDA table ($\beta = 1.10; \chi^2(1) = 10.46, p < .01$) and for the Choices Logo ($\beta = 0.78; \chi^2(1) = 6.44, p < .05$). For task condition, labels were mentioned less often after a preference task than after a health task ($\beta = -1.19; \chi^2(1) = 11.99, p < .01$) or a low saturated fat task ($\beta = -1.34; \chi^2(1) = 14.30, p < .001$), again in line with results from prior research (Van Herpen & Van Trijp 2011). The health and low saturated fat conditions did not significantly differ in their effect on label mentioning.

3.2.7. Perceived healthfulness

Healthfulness ratings were subjected to a repeated measures ANOVA, with pizza type (more, medium, less healthful) as the repeated measures factor and labeling scheme, task, and country as between subject factors. Results showed a main effect of pizza type ($F(2, 718) = 240.63, p < .001$) with a high mean rating for the more healthful pizzas ($M = 4.90$), an intermediate rating for the medium healthful pizzas ($M = 3.73$), and a low rating for the less healthful pizzas ($M = 3.50$). There was also a significant pizza type \times label ($F(6, 718) = 12.18, p < .001$) interaction, but the pizza type \times task interaction was not significant. We did find a significant main effect of label ($F(3, 359) = 4.90, p < .01$) and a label \times task interaction ($F(6, 359) = 3.01, p < .01$; for the preference task only, pizzas generally received higher ratings in the Logo condition than in the other labeling conditions. None of the other main effects or interactions was significant. Thus, country had no significant effect on the perceived healthfulness of different pizzas, nor did it interact with the labeling schemes.

Fig. 1 provides a graphical display of the pizza type \times label interaction. To examine this interaction further, we focused on each of the three pizza types (more, medium, less healthful) separately. For more healthful pizzas, there was a significant effect of labeling scheme ($F(3, 359) = 10.11, p < .001$) and post hoc tests (LSD) indicated that the no label condition was significantly different from all other conditions ($ps < .05$), with no significant differences between the labeling schemes themselves ($ps > .05$). In other words, all labeling schemes were equally able to raise healthfulness perceptions for more healthful pizzas, which is in line with Hypothesis 6a. For less healthful pizzas, healthfulness evaluations were affected by labeling scheme ($F(3, 359) = 12.39, p < .001$). Post-hoc tests revealed that the no label condition and the Logo condition did not significantly differ ($p = .7$), whereas both the MTL label and the GDA label significantly decreased healthfulness ratings for these pizzas compared to the control condition ($ps < .001$), thus lending support for Hypothesis 7a. For medium healthful pizzas, there was a marginal effect of labeling scheme ($F(3, 359) = 2.29, p = .08$). Post-hoc tests revealed that the no label condition and the Logo condition did not significantly differ ($p > .4$), whereas the MTL and GDA labels led to significantly lower ratings than the no label control condition ($ps < .05$). Thus, whereas the MTL and GDA label decreased healthfulness perceptions for medium healthful pizzas, the Logo did not.

3.2.8. Product attractiveness

Repeated measures ANOVA was also applied on the attractiveness ratings. Results were in line with those found for healthfulness perceptions. Again, there was a main effect of pizza type ($F(2, 718) = 138.36, p < .001$) with a high mean rating for the more healthful pizzas ($M = 5.04$), an intermediate rating for the medium healthful pizzas ($M = 4.03$), and a low rating for the less healthful pizzas ($M = 3.69$). There were also significant pizza type \times task ($F(4, 718) = 3.18, p < .05$) and pizza type \times label ($F(6, 718) = 2.95, p < .01$) interactions. None of the other main effects or interactions

Table 4
Product choice, attractiveness, and perceived healthfulness.

Sample size	Labeling scheme							
	No label		Choices Logo		MTL label		GDA table	
	UK	NL	UK	NL	UK	NL	UK	NL
Choice ^a	23	23	30	37	40	39	34	36
More healthful	53.5	46.0	62.5	72.5	81.6	83.0	73.9	73.5
Medium	25.6	20.0	12.5	13.7	10.2	8.5	13.0	10.2
Less healthful	20.9	34.0	25.0	13.7	8.2	8.5	13.0	16.3
Label mentioned ^b	0	0	31.3	39.2	46.9	57.4	30.4	30.6
Attractiveness								
More healthful	4.69	4.75	5.15	5.15	4.97	5.42	5.05	5.10
Medium	4.15	4.42	4.06	4.23	3.83	3.94	3.80	3.80
Less healthful	3.74	3.79	4.00	3.63	3.50	3.84	3.70	3.34
Healthfulness								
More healthful	4.24	4.38	5.39	5.13	5.02	5.11	5.09	4.83
Medium	3.80	4.02	3.90	3.73	3.49	3.71	3.63	3.58
Less healthful	3.94	3.59	3.85	3.77	2.96	3.22	3.22	3.04

^a In percentages.

^b In answer to an open question on choice motivation, in percentages.

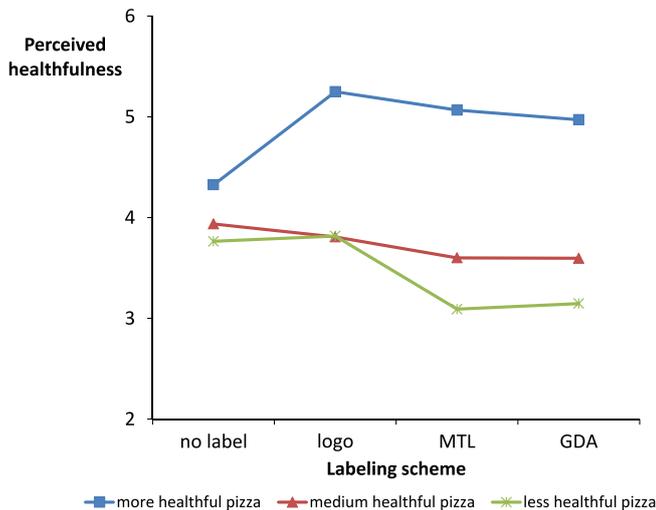


Fig. 1. Perceived healthfulness of healthful, medium, and unhealthy pizzas as a function of labeling scheme.

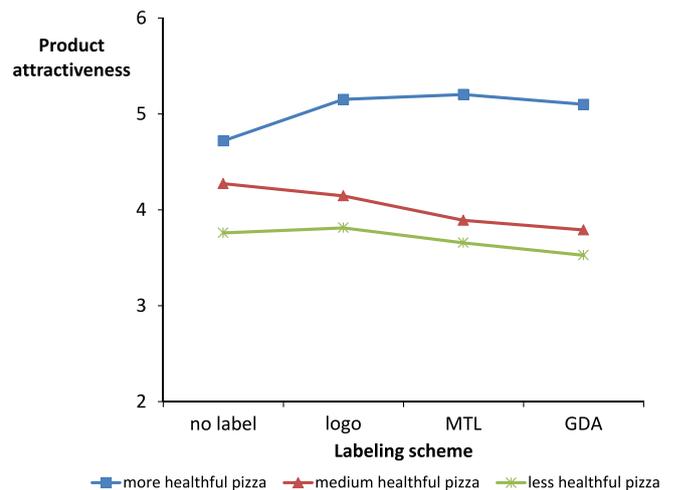


Fig. 2. Attractiveness for healthful, medium, and unhealthy pizzas as a function of labeling scheme.

were significant. Thus, in line with results for perceived healthfulness, country had no significant effect on the attractiveness of different pizzas, nor did it interact with the labeling schemes.

Fig. 2 provides a graphical display of the pizza type \times label interaction for attractiveness ratings. To examine this interaction further, we again focused on each of the three pizza types (more, medium, less healthful) separately. For more healthful pizzas, in line with the results for perceived healthfulness, there was a significant effect of labeling scheme ($F(3, 359) = 3.53, p < .05$) and post hoc tests (LSD) indicated that the no label condition was again significantly different from all other conditions ($ps < .05$), with no significant differences between the labeling schemes themselves ($ps > .4$). In other words, all labeling schemes were equally able to raise attractiveness for more healthful pizzas, which is in line with Hypothesis 6b. For less healthful pizzas there was no significant main effect of labeling scheme ($F(3, 359) = 1.04, ns$), and as a result Hypothesis 7b was not supported. These pizzas had relatively low attractiveness scores in the control condition, and the labels did not further decrease these attractiveness ratings. Finally, for medium healthful pizzas, there was also a significant main

effect of labeling scheme ($F(3, 359) = 3.34, p < .05$). Post-hoc tests showed the same pattern of results as we found for perceived healthfulness: the no label condition and the Logo condition did not significantly differ ($p > .3$), whereas the MTL and GDA labels led to significantly lower ratings than the no label control condition ($ps < .05$). Thus, whereas the MTL and GDA label led participants to prefer medium healthful pizzas less, the logo did not.

The pizza type \times task interaction reflected that task only had a significant effect on attractiveness for more healthful pizzas ($F(2, 359) = 5.12, p < .01$) and not for the medium or less healthful pizzas. For more healthful pizzas, a preference task led to lower attractiveness ratings ($M = 4.79$) than either the health task ($M = 5.20$) or the low fat task ($M = 5.13$), which did not significantly differ.

4. Discussion

The present study explores the effectiveness of alternative nutrition labeling schemes in stimulating (healthfulness) perception and the healthful choice among consumers. It shows that familiarity with particular labeling schemes mediates label

evaluations, but not healthful choices. This implies that consumer evaluations of labeling schemes need to be interpreted with care, as they may be specific to the market where they are collected and findings may not generalize to markets with different label familiarity patterns. However, such biased evaluation of labels, by familiarity, does not translate to the use of labels in choice.

All labels are found to be equally effective in enhancing healthful choice. In this respect, the present study confirms previous findings that consumers' self-reports on label evaluation and usage intentions are not appropriate measures for the effectiveness of nutrition labeling in product choices. Non-directive labels (nutrition tables and monochrome GDA) are evaluated most positively, but other labels are equally (this study) or even more (e.g., Van Herpen & Van Trijp 2011) effective in enhancing the healthful choice. One might argue that the current comparison has been harsh to the Choices Logo for two reasons. First, salience within the visual field is an important determinant of consumer attention to nutritional labels (Bialkova & Van Trijp, 2010) and this puts the Choices Logo to a disadvantage as it is smaller than the other labels, which are also present on all products rather than only the more healthful options. Also, in the present study, the GDA and MTL labels, but not the Choices Logo, were augmented with calorie information, which is known to be an important cue to consumers in identifying the more healthful option (Van Kleef et al., 2008).

Another finding is that different labeling schemes increase transparency within an assortment in different ways. In terms of attractiveness and healthfulness perception, all labeling schemes reward the more healthful options within the assortment. Both the MTL and the GDA label (but not the Choices Logo) additionally "punish" the less healthful options in reducing their healthfulness perception, but without affecting their attractiveness ratings. For the medium healthful pizzas, MTL and GDA (but again not the Choices Logo) decrease both healthfulness perceptions and attractiveness ratings. This suggests that in terms of attractiveness the MTL and GDA labels particularly impact on the medium healthful pizzas.

Together these findings provide an important step in elucidating the underlying cognitive processes on the working of front-of-pack nutrition labels. Integrating the findings with research on other on-package information would be promising to further examine the cognitive processes involved. For instance, the depth of processing of on-package information ranges from a mere glance, to a reliance on what is familiar, to the construction of new associations (Hoogland, de Boer, & Boersema, 2007), and different types of on-package information may be more effective depending on this depth of processing, as future research may examine.

Managerially and in terms of policy, the present study has several implications. First, it shows that self-reports of consumers should be considered with care. Self-reports and stated usage intentions seem to be affected by a priori familiarity much more than usage and effect of the labeling schemes. Additionally, self-report based studies would particularly add insight when conducted in a cross-country context as the generalization of single country results is questionable. There is a lack of comparative studies in previous research and it cannot be ruled out that much of the current policy debate is at least partially affected by country bias. This calls for cross-cultural comparative research where self-reported evaluations are augmented with actual choice data in response to different labeling schemes.

Although our study provides a next step in advancing insight into the effectiveness of nutrition labels, it is not without its limitations. First, the present study addresses only two countries (UK and the Netherlands) and for pan-European policy implications it would need to be extended to more and more culturally diverse

countries across Europe. Further, the present study used specific implementations of nutrition labeling. Although the most prominent labeling schemes (Choices Logo, MTL, and MC-GDA) were included, other nutrition labels (such as Health Tick, Swedish Key Hole, and CC-CGA) could also be further examined. Each of these labels could appear in multiple format executions. In the present study we selected a context in which MTL and GDA were accompanied with a calorie label, whereas the Choices Logo was not. Other executions are feasible also to further disentangle the nutrition labeling effects on product choice due to the nutrition label format per se versus those due to the addition of calorie labeling. Another extension left for future research is the examination of consequential choices, as the choices that participants made in our study were hypothetical. Additionally, healthfulness and attractiveness ratings were provided based on memory of the products to prevent participants from focusing on information they would not naturally consider. As memory can be inaccurate, results might have been stronger if products would have remained visible during ratings.

Finally, the present study was based on University students in the UK and the Netherlands, rather than "mainstream" consumers in those two markets. Consumers with a higher education, such as these students, generally understand nutrition labels relatively well (Drichoutis et al., 2006), and participants also received information about the labeling schemes. Still, results show that intentions to use labels are affected by label familiarity, indicating that a good understanding of all labels is not enough to overcome potential biases due to label familiarity, an important caveat for research into food labeling. Nonetheless, it would be important to explore the generalization of the present findings to more mainstream consumer markets in the UK, the Netherlands and beyond.

5. Conclusion

The present study addressed a gap in existing nutrition labeling research, namely that existing research has largely been based on consumers' self-reports and often has taken a country specific approach. It shows that self-reported evaluation and usage intentions of labeling schemes are not an accurate predictor of their effect on healthful product choices. Both in the UK (where MTL and GDA labels prevail) and in the Netherlands (where Choices Logo and GDA are most widely used), all three label types are effective in enhancing the healthful product choice, despite that a priori familiarity with the labeling schemes affects consumers self-reported evaluation and usage intention. The results further confirm that in terms of healthfulness perception all labels "reward" the more healthful options within the assortment, but only the MTL and GDA also disqualify other options. In particular, these latter labels appear to decrease attractiveness for medium healthful product options, rather than less healthful options.

Acknowledgements

The authors gratefully acknowledge the European Community financial contribution under the Seventh Framework Program for Research, Technological Development and Demonstration Activities, for the Small Collaborative Project FLABEL (Contract No. 211905). The content of the paper reflects only the views of the authors; the European Commission is not liable for any use that may be made of the information contained in this paper.

The authors are indebted to Ramon Roos, Charo Hodgkins, Monique Raats, Ashley Guinn, and Kris Smith for their help in data collection, and to Stefan Storcksdieck genannt Bonsmann for feedback on an earlier version of this paper.

Appendix A

The assortment of pizzas with different labeling schemes.

Choices logo



MTL label



Monochrome
GDA label



Note: Control condition contained the same pizzas without any labels

Appendix B

Information explaining the different labeling formats.

Packaged foods often contain labels or symbols which can help you make your food choice. In this information leaflet you will find information about different food labels that you may encounter.

Multiple Traffic Light

Each Traffic Light shows if the product contains a low, medium or high amount of sugar, fat, saturated fat or salt per 100 grams. The colour green indicates that the product contains a low amount of the nutrient, the colour amber indicates an average amount and the colour red indicates a high amount.



Health Tick

The Health Tick can be found on products that are a healthy choice within their product category. The criteria for the logo differ for each product category. Products with this logo contain a lower amount of sugar, fat, saturated fat and salt per 100 grams.



GDA Table

The GDA Table shows detailed information about the nutrient contents of a product. The table tells you the amount of sugar, fat, saturated fat and salt in 100 grams of the product. It also gives you the percentage of the Guideline Daily Amount.

CAL 203 per 100 gram	Sugars	Fat	Sat Fat	Salt
	2.7	3.4	1.4	0.31
	3%	5%	7%	13%

References

- Andrews, J. C., Burton, S., & Kees, J. (2011). Is simpler always better? Consumer evaluations of front-of-package nutrition symbols. *Journal of Public Policy & Marketing*, 30, 175–190.
- Balcombe, K., Fraser, I., & Di Falco, S. (2010). Traffic lights and food choice. A choice experiment examining the relationship between nutritional food labels and price. *Food Policy*, 35, 211–220.
- Bialkova, S., & Van Trijp, H. (2010). What determines consumer attention to nutrition labels? *Food Quality and Preference*, 21, 1042–1051.
- Bialkova, S., & Van Trijp, H. (2011). An efficient methodology for assessing attention to and effect of nutrition information displayed front-of-pack. *Food Quality and Preference*, 22, 592–601.
- Borgmeier, I., & Westenhoefer, J. (2009). Impact of different food label formats on healthiness evaluation and food choice of consumers: A randomized-controlled study. *BMC Public Health*, 9, 184.
- Bussell, G. (2005). Nutritional profiling vs guideline daily amounts as a means of helping consumers make appropriate food choices. *Nutrition & Food Science*, 35, 337–343.
- Drichoutis, A. C., Lazaridis, P., & Nayga Jr., R. M. (2006). Consumers' use of nutritional labels: A review of research studies and issues. *Academy of Marketing Science Review*, 9. Available at: <http://www.amsreview.org/articles/drichoutis09-2006.pdf>.
- Feunekes, G. I. J., Gortemaker, I. A., Willems, A. A., Lion, R., & Van den Kommer, M. (2008). Front-of-pack nutrition labelling: Testing effectiveness of different nutrition labelling formats front-of-pack in four European countries. *Appetite*, 50, 57–70.
- Gorton, D., Mhurchu, C. N., Chen, M.-H., & Dixon, R. (2009). Nutrition labels: A survey of use, understanding and preferences among ethnically diverse shoppers in New Zealand. *Public Health Nutrition*, 12, 1359–1365.
- Grunert, K. G., Fernández-Celemin, L., Wills, J. M., Storcksdieck genannt Bonsmann, S., & Nureeva, L. (2010a). Use and understanding of nutrition information on food labels in six European countries. *Journal of Public Health*, 18, 261–277.
- Grunert, K. G., & Wills, J. M. (2007). A review of European research on consumer response to nutrition information on food labels. *Journal of Public Health*, 15, 385–399.
- Grunert, K. G., Wills, J. M., & Fernández-Celemin, L. (2010b). Nutrition knowledge, and use and understanding of nutrition information on food labels among consumers in the UK. *Appetite*, 55, 177–189.
- Halliday, J. (2009). Campaign aims to stop GDA inclusion in labelling law. *Food Navigator.com, William Reed Business Media*. Available from: <http://www.foodnavigator.com/Legislation/Campaign-aims-to-stop-GDA-inclusion-in-labelling-law>.
- Hieke, S., Wilczynski, P., & Schwaiger, M. (2011). Color me in. An empirical study on the traffic light signposting system. *Association for Marketing & Health Care Research (AMHCR) Conference Proceedings*.
- Hodgkins, C., Barnett, J., Wasowicz-Kirylo G., Stysko-Kunkowska, M., Gulcan, Y., Kustepeli, Y. et al. (2011). Understanding how consumers categorise nutritional labels: A consumer derived typology for front-of-pack nutrition labeling. Working paper.
- Hoogland, C. T., de Boer, J., & Boersema, J. J. (2007). Food and sustainability: Do consumers recognize, understand and value on-package information on production standards? *Appetite*, 49, 47–57.
- Jones, G., & Richardson, M. (2007). An objective examination of consumer perception of nutrition information based on healthiness ratings and eye movements. *Public Health Nutrition*, 10, 238–244.
- Kelly, B., Hughes, C., Chapman, K., Louie, J. C.-Y., Dixon, H., Crawford, J., et al. (2009). Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market. *Health Promotion International*, 24, 120–129.
- Kozup, J. C., Creyer, E. H., & Burton, S. (2003). Making healthful food choices: The influence of health claims and nutrition information on consumers' evaluations of packaged food products and restaurant menu items. *Journal of Marketing*, 67, 19–34.
- Levy, A. S., Fein, S. B., & Schucker, R. E. (1992). More effective nutrition label formats are not necessarily preferred. *Journal of the American Dietetic Association*, 92, 1230–1234.
- Maubach, N., & Hoek, J. (2008). The effect of alternative nutrition information formats on consumers' evaluations of a children's breakfast cereal. *Partnerships, Proof and Practice – International Nonprofit and Social Marketing Conference – Proceedings*. Paper 1. Available from: <http://ro.uow.edu.au/insm08/1>.
- Maubach, N., Hoek, J., Healey, B., Gendall, P., & Hedderley, D. (2009). Motivation, ability and the influence of nutrition information formats. *ANZMAC 2009 Conference*.
- Maubach, N., & Hoek, J. (2010). A qualitative study of New Zealand parents' views on front-of-pack nutrition labels. *Nutrition & Dietetics*, 67, 90–96.
- Mhurchu, C., & Gorton, D. (2007). Nutrition labels and claims in New Zealand and Australia: A review of use and understanding. *Australian and New Zealand Journal of Public Health*, 31, 105–112.
- Möser, A., Hoefkens, C., Van Camp, J., & Verbeke, W. (2010). Simplified nutrient labelling: Consumers' perception in Germany and Belgium. *Journal of Consumer Protection and Food Safety*, 5, 169–180.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89, 852–863.
- Sacks, G., Rayner, M., & Swinburn, B. (2009). Impact of front-of-pack 'traffic-light' nutrition labelling on consumer food purchases in the UK. *Health Promotion International*, 24, 344–352.
- Scott, V., & Worsley, A. F. (1994). Ticks, claims, tables and food groups: A comparison for nutrition labelling. *Health Promotion International*, 9, 27–37.
- Storcksdieck genannt Bonsmann, S., Fernández Celemin, L., Larrañaga, A., Egger, S., Wills, J. M., Hodgkins, C., et al. (2010). Penetration of nutrition on food labels

- across the EU-27 plus Turkey. *European Journal of Clinical Nutrition*, 64, 1379–1385.
- Storcksdieck genannt Bonsmann, S., Fernández Celemin, L., & Grunert, K. G. (2010). Food labelling to advance better education for life. *European Journal of Clinical Nutrition*, 64, S14–S19.
- Van Herpen, E., & Van Trijp, H. (2011). Front-of-pack nutrition labels. Their effect on attention and choices when consumers have varying goals and time constraints. *Appetite*, 57, 148–160.
- Van Kleef, E., Van Trijp, H., Paeps, F., & Fernández-Celemin, L. (2008). Consumer preferences for front-of-pack calories labeling. *Public Health Nutrition*, 11, 203–213.
- Visschers, V. H. M., Hess, R., & Siegrist, M. (2010). Health motivation and product design determine consumers' visual attention to nutrition information on food products. *Public Health Nutrition*, 13, 1099–1106.
- Vyth, E. L., Steenhuis, I. H. M., Mallant, S. F., Mol, Z. L., Brug, J., Temminghoff, M., et al. (2009). A front-of-pack nutrition logo: A quantitative and qualitative process evaluation in the Netherlands. *Journal of Health Communication*, 14, 631–645.
- White, J., Thomson, G., & Signal, L. (2010). Front-of-pack nutrition labelling: Where to now? *The New Zealand Medical Journal*, 123, 12–16.