Shopper's behavioural responses to 'front-of-pack' nutrition logo formats: GDA Diet-Logo vs. 6 alternative Choice-Logos

Laurent Muller (INRA UMR 1215 GAEL),

Bernard Ruffieux (Grenoble Institute of Technology)

Abstract: A Framed Field Experiment was implemented in France in order to compare the relative behavioural responses and the induced nutritional effectiveness of seven front-of-pack logo formats: GDA (the science-based generated 'diet logo'), and six behavioural-based 'choice logos' such as keyhole symbol or traffic lights. Choice logo formats are distinguished by 3 criteria: the unit ranked, the reference set and the scope of colours. Results show that some choice logos perform better that GDA in improving the nutritional quality of shopper baskets. A choice-logo would induce better behavioural responses if designed on an aggregated and multi-coloured indicator.

Key words: food consumption, nutritional labelling, food policy, framed field experiment, behavioural economics

JEL codes: D120, D180, C910, C930.

Introduction

In order to reverse the trends of obesity and other food related chronic diseases, should public deciders rely solely on the capacity of consumers to make the right decisions when they are shopping for food? A front-of-pack labelling system may potentially change consumers' behaviour toward healthier diets. However contrasted logo formats featuring various options

are available for such a system. This study aims at measuring the behavioural consumer response to such alternatives in order to identify the best possible options.

Standard economic theory assumes that individuals maximize their own well being subject to a set of constraints. In food contexts, this means that, given the information available and the limits on income or time (to prepare food for instance), consumers are making their best possible food choices and, therefore, are properly arbitrating between short-term gains of sensory pleasure and longer-term gains of health and wellness. From there, the task of public deciders should only consist in fixing externalities (*e.g.* social cost of obesity) by altering for example relative prices through taxes and subsidies and in providing the necessary information – as it is not directly accessible through the consumption of the product – that will allow the consumers to actually make their best choices.

Considerable efforts have been made to fully inform consumers. For instance, the back of most food packs contains, in addition to the list of ingredients, a nutrition panel, i.e. information on content in each nutrient, completed or not with daily values or Guideline Daily Amounts (GDA). With such tools, optimizing diets is made possible yet remains highly complex: the maximization program would include thousands of products and dozens of nutrients with uncertain inter-temporal outcomes. At best, sophisticated and concerned consumers will strive to construct an ideal diet based on the GDA recommendations. In reality, shoppers are distracted, pressed for time and thus pay little attention to each of their multiple food decisions.

If we accept that food-related behaviours are not always derived from rational decision-makings, public deciders may consider nudging consumers towards healthier diets by presenting health information in ways that take in account shoppers' limited rationality. The first objective is to catch consumer attention: salient logos may be displayed front-of-pack

and visible at a glance. The second objective is to simplify consumer choice towards healthier option by easing products substitutions. Accordingly, logos ought to be prescriptive; they would give direct and clear hints about the nutritional quality of the product compared to others without referring to the global diet. Such logos are already in use like the British *shopping tips* or *traffic light* or the Scandinavian *Keyhole Symbol*. Both differ from GDA by systematically providing a normative assessment. They also fundamentally differ from each other: while *Traffic lights* notify the level of each nutrient relative to all products, the keyhole symbol reward the healthiest options among the products of the same category. Which one is the best?

In a framed field experiment, this study compares the behavioural responses to seven front-of-pack logo formats: one is a 'diet logo' Guideline Daily Amounts (GDA); six are 'choice logos'. While GDA is science-based and requires a global-diet heuristic, the choice logos are behavioural-based and require simpler product comparison heuristics. The six formats are differentiated by 3 criteria: the unit ranked, the reference set and the scope of colours. We first show that front-of-pack logos significantly improve the nutritional quality of diets. Second, we find that different logo formats generate different behavioural responses. Finally, we rank logo formats according to the extent of nutritional improvement of consumers shopping baskets: some choice logos perform better than GDA; an aggregate indicator induces better responses; a multicolour indicator is globally better but causes nutritional deterioration for more individuals; and category logos are equivalent to transversal logos although they trigger radically different behavioural trajectories. These results highlight the limited attention, the quick decision fatigue and trade-off aversion, but also the compliance of consumers.

Background

It is now well documented that, over the past few decades, trends in food consumption have been associated with an increase in food-related chronic diseases such as obesity, certain cancers, coronary heart disease, diabetes, and gut diseases (WHO 2003; Cutler, Glaeser and Shapiro 2003; Loureiro and Nayga 2005; Bleich *et al*. 2008; Etilé 2011). Therefore, improving the nutritional quality of food consumption is now among the top priorities of public food authorities all over the world (WHO 2003).

The basic components of the toolkit for such a nutritional policy is well accepted: change in the food environment, products quality food standards regulation (Swinnen and Vandemoortele 2011), restrictions on advertising (Chou, Rashad and Grossman 2008), public campaigns and education to provide better nutritional knowledge (Etilé 2011), price policies using taxes and subsidies to counterbalance long-term price trends (Huang 1996; Allais, Bertail and Nichèle 2010; Caraher and Cowburn 2005), product nutritional labelling (Drichoutis, Nayga and Lazaritis 2011). New alternative policies include the use of behavioural economics (Just, 2006; Just and Payne 2009) and nudge (Campbell-Arvai, Arvai and Kalof 2012; Thaler and Sunstein, 2008).

Since the 1970's, when official reports began to recommend nutritional guidelines for consumers, food packaging labelling has been recognized as a political tool that might be used to encourage consumers to adopt a healthier diet, and now most political authorities recognize this labelling policy as both a key and consensual instrument (Drichoutis, Lazaridis and Nayga 2006; Grunert and Wills 2007). We focus here on such labelling policies.

Back-of-pack panels and rational optimization

The objective of back-of-pack panels is to define accurate and reliable information contents. Biologists, nutritionists and medical scientists are therefore at the forefront in order to produce highly credible science-based 'positive' knowledge. The content of such panels vary from one country to another. In the US, information is displayed in listing global calories content, and gives useful information in nutrients content, listing nutrients that should be limited – sodium, fat, sugar, cholesterol, etc. – and nutrients that should be favoured, such as vitamins, calcium, iron, etc. For each nutrient, information is given either per 100g or per serving. A simple back-of-pack panel may display more than 30 figures in absolute terms or percentages. It may even contain much more.

Let us consider the rational behaviour that lies behind such panels. Suppose a consumer considering buying a pizza. To decide appropriately, she needs to think globally, to identify the marginal impact of such a choice on her daily diet by analysing, nutrient by nutrient, the intake of the pizza. She has to integrate in her decision process the content of the pizza (low content in sugar and high content in salt and fat) and the content of all the other products included in her diet and consider what could complement or substitute the pizza (adding a soda? reducing crisps?). Even if we restrict the nutrition optimization program to the three major nutrients to limit (sugar, fat and salt) the computation skill is certainly beyond reach for any consumer. In order to ease decision fatigue but remaining in the same heuristic, *Daily Values* have been implemented. Daily values refer to a standardized 2000 calories diet of a typical adult, suggesting the recommended daily intake for each nutrient. Then for each product consumers know the daily percentage of each nutrient provided by that product based on an ideal diet. Conforming to this, consumers are informed about the daily percentage of each nutrient provided by a given product and may choose accordingly.

Heuristics and front-of-pack logos

In addition to the high complexity of food optimization, nutrition is only one out of many other characteristics such as price, taste, safety, sustainability, etc. Due to the fact that the health impact is hidden or delayed after immediate use, most of these characteristics may be regarded as more straightforward and thus even prioritized. Furthermore, supermarkets are not precisely the best place for slow and rational behaviour. Food shoppers are usually in a hurry, they are unmindful of health; attracted by colours, trademarks, and the marketing environment (Wansink 2004). It has been estimated that people make an average of 200 to 300 decisions regarding food consumption in any given day (Wansink and Sobal 2007). Therefore new heuristics may be required to shorten and ease decision-making. Heuristics may also lead to well-documented biases (DellaVigna 2009). It is in the hand of public deciders to take advantage of these biases to efficiently guide consumers towards healthier diets. Front-of-pack logos can become a helpful tool for food policy. Implemented in Sweden since 1989 and later spread over Scandinavia, the keyhole symbol identifies as 'green', with a unique aggregated index, products that are nutritionally best within each food category. UK shopping tips knows as traffic lights, expend the scoop of colours to green, yellow and red to identify what's good, average, or bad. Shopping tips refer to each nutrient but a single can also refers to the product as a whole. US key icons after the "2010 First Lady Michelle Obama initiative" move GDA from back to front-of-pack ("facts up front!")[1].

As described in the precedent paragraph, GDA refers to an ideal diet and thus requires a global heuristic. On the other hand, traffic lights and keyhole symbols directly relate products. Colours and icons have been added and most digits removed. Given that attention, speed (ready-access), perception, saliency, the link between the nature of information and decision-

making, etc. are important issues, behavioural economics can make a valuable contribution. The present paper is intended to be a step in this direction.

Behavioural Proposals and logo formats selected for experimental testing

In this section we explicit six proposals based on behavioural economics theoretical body and previous experimental results. They will come in theoretical support for our selection of logo formats that will include simplified variants of the most significant formats presently at use in actual public policies, such as the UK traffic light, the Scandinavian keyhole symbol, and the US GDA. Each format is distinguished by three criteria. A criterion is an attribute of a logo format that takes alternative options. In our selection, we only consider two extreme options per criterion. We will also only consider the density (g per 100g) of the following three nutrients: salt, free sugar and saturated fatty acid. These nutrients are consensually recognized among nutritionists as the nutrients to be limited. Finally, the following sorting rule will be used: the best 1/3 will be green, the worst 1/3 will be red and the remaining 1/3 will be colourless.

Proposal 1. A front-of-pack logo may make a difference, whatever the format

By catching attention and stressing nutritional concern, front-of pack logos may significantly improve the nutritional quality of the consumer basket. A typical consumer is making fast shopping decisions with poor consideration to nutritional issues. *Attention* plays a key role for a better consideration of nutritional quality in consumer's preferences. Once the attention is captured, the weight of the health component in food choice should increases relatively to other food attributes. If consumers are not totally ignorant of nutritional issues, they may broadly identify what is nutritionally good or bad for him and somewhat improve his diet. In

order to be fully convinced by the proposal, let us follow a small *deduction ad absurdum*: Imagine that if consumers were concerned (and ready to invest in efforts) but ignorant of nutritional issues, back-of-pack panels would then become efficient. We know this is not the case.

Proposal 2. Behavioural responses may differ according to the logo format

Different formats may induce different behavioural responses in magnitude and quality. As we will see in proposals 4, 5 and 6, some criteria options may require more cognitive efforts or leave the consumer with more tricky trade-offs to resolve. In such cases, we assume that less change will be generated. In other cases, different options may direct consumers towards different type of product substitutions. As a result, the nutritional impacts induced may differ.

Proposal 3. GDA is nicely designed to assist extremely rational behaviours

Due to *limited attention*, *limited computing capability* and *lack of global thinking*, GDA may not be the most efficient logo format. GDA provides consumers the numerical tools to construct a perfect diet, *i.e.* one that fits the daily-recommended intake exactly for each nutrient. It does not say anything about *how* to do it and it is up to the consumer to resolve an optimization problem under constraints. While *Homo Oeconomicus* would eventually end up with an ideal diet, behavioural men do not compute all information available and most decisions are taken on the basis of inward and local thinking (Ariely 2009).

Proposal 4. Unit graded: Simplicity may pay

As we saw about GDA, the more information the better for those who are able to process it. Analytically similar to GDA, a logo may evaluate each nutrient. Another possibility is to evaluate the whole product with a unique global indicator that aggregates the nutrients contents. Although consumers (for example through focus groups) declare a desire for precise

nutritional data on products, a quick and easy-to-check logo, avoiding difficult trade-offs reduces efforts and thus stimulates more changes. On the other hand, an aggregated logo may also induce useless efforts, or counterproductive efforts if hidden trade-offs appear (*e.g.* negative correlations among nutrients)[2].

This leads to the design of our first criterion among the three defining a choice logo format: namely the selection of the *unit graded*. We propose to test two options for this criterion: option N and option P. Option N will be when each nutrient of a given product is graded as such. Therefore, the N logo format will include three qualifications (colour patches in our experiment), one for each of our three nutrients. Option P will be when the unit graded is the product. One equation to aggregate the three nutrient grades in one aggregated grade will be used. Nutritional scientists generally accept such a *modus operandi* therefore one of them, the LIM indicator (Darmon et al. 2009), has been selected.

Proposal P5. Reference set: substituting among the same category is less costly in effort and less rewarding in nutrition

Here, the question is whether it would be more efficient to rank products in reference to the whole food supply or to a category of products (a category comprises close substitutes). The reference set induces different substitution responses, though their nutritional impacts are uncertain. Intra-category substitutions may require less effort for consumers (crisps vs. light crisps). On the other hand, the nutritional gain may be less important than inter-category substitutions (crisps *vs.* radish)[3].

Thus, our criterion 2 affects the reference set applied for sorting the products into our three grading nutritional qualities: best, average, and inferior. Two options are possible: option C for '*Category set*'; and option O for '*Overall set*'. With option C, each product is graded according to its product category. For example, a 'croissant' is ranked in the category

‘Viennese pastries and cakes’. With the O option, a product is ranked referring to all food products in our experimental E-Shop.

Proposal P6. Scope of colour: Red and green is more efficient than only green

The final issue is to decide whether the logo format will identify only good products to the consumer (green) or both good and bad products (red and green). We have already stated in P3 and P4 that more information gives users who are able to process it more openings to improve their food basket. Therefore, extending the display to red indicators, in addition to green ones, improves the efficiency. Additionally, a red indicator, with its off-putting effect, may make a strong impression on participants. It may change not only their information set but also their preferences. This change may be *emotional* and therefore transitory, yet significant. *Velleity* furthermore plays a big part, especially when food and nutrition are concerned. A balanced diet is always postponed to tomorrow (Ariely 2011). With an emotional red patch, some substitutions or some relinquishments may be less subject to procrastination than others. Note also that some participants may value *compliance* while others refuse it (Kelman 1958). The scope of colour is important in terms of policy design because a green format may be introduced as ‘voluntary’ and a traffic light format as ‘mandatory’, with potentially hostile reactions on the supply side. Is it worth trying?

Accordingly, our criterion 3 refers to the display extent with which information is posted on the logo. The two options are: option G for ‘*Green*’ and option T for ‘*Traffic-light*’. In relation to option T, only one third of the products, those nutritionally sorted as best, are signalled. A green indicator appears on the logo of such products. Note that, with option N, a product’s logo may possibly exhibit either one, two or three green patches, *i.e.* one for each nutrient for which the product is among the best third, referring to the relevant reference set.

Seven logo formats in competition in the Lab

To sum up, each logo is defined as a combination of three options, one for each criterion. The first criterion, the *unit ranked*, may refer to the whole *product* (P) or to each *nutrient* (N); the second criterion, the *reference set*, may refer to the *category* subset (C) or *overall* product set (O); the third criterion, the *scope of signalling*, may take the option *green* only with only the best products marked (G) or *traffic light* where best and worst products are marked (T). Six of the eight possible formats that this system generates have been selected; the other two are too complex to be promising[4]. Thus a selection of six logo formats with which to compare GDA is proposed (See table 1). Salt, free sugar and saturated fatty acid, the three nutrients for which consumption needs to be reduced[5], have been selected. Products or nutrients are sorted into three equal groups: one-third best; one-third average; one-third inferior[6].

Experimental Protocol

Using a framed field experiment with consequential choices (Harrison, List 2004), a laboratory e-shopping environment was mocked-up to observe shoppers' actual behavioural response to seven alternative front-of-pack nutrition logo formats. This section first describes the e-Shop implemented in the lab (products, food categories, catalogues and graphical user interface). Then the architecture of the experiment, the incentive mechanisms, the subjects' recruitment and the session organization will be presented. Finally, the principles of data analysis used in the results' section will be described.

Implementing an E-Shop in the Lab

The experimental e-shop includes a total of 273 food products[7] in 35 familiar food categories[8]. Defining the list of food items available in the e-shop and sorting these items into categories

is a key feature as such choices are critical for the relevance and generalization of the results. Unfortunately and unavoidably, it contains *ad hoc* features. Categories are particularly crucial for the C options where products are graded in reference to the category. In this experiment, categories correspond to the usual classification used in self-service grocery stores in France. It also fits the standard classification proposed by OQALI[9]. Each category includes either six or nine different products (with the exception of two categories that comprise twelve) in order to ease our ranking configuration (1/3 best, 1/3 worst and 1/3 neutral). Food items have been chosen among the most frequently bought products in France in each category, so as to model the existing range of nutritional quality in each category. Products were proposed at current outside market prices. Posted prices had been observed in a local supermarket at the time of the sessions. Participants were aware of that.

A paper catalogue is available for each subject during the experiment. It is a 35-page A4 format colour booklet. Each page includes the items of a category. The category is named at the top of the page. On each page a coloured front-of-pack picture of each product with the name of the product, the price per unit and a bar code appear. Next to a personal computer each participant has an individual easy-to-use bar code reader. By reading any product code with the code reader, the participant makes this product pop up on the computer screen. The participant may then use the computer keyboard to buy one or more units of the selected item. On the right side of the screen the work-in-progress basket appears. It includes the name of items already selected, the price of each item and the total amount already spent. Any selected item may easily be removed from the basket during the shopping stage[10].

Architecture of the Experiment and Treatments

Participants are first invited to shop for food for the household. A *food basket* is defined as all food items bought from the e-shop to feed household members over two days following the

experiment. Each session is devoted to one of the seven treatments (one per logo format) and comprises 3 stages. In stage 1, subjects are asked to fill their food basket in the absence of logos. We called the stage 1 basket the *reference basket*. At this point, participants are unaware of the purpose of the experiment and that further stages are to come. In stage 2, one logo option is introduced and explained to the participants. Logos are then applied exhaustively to the 273 products and are visible online and in new catalogues. Everything else remains unchanged. Participants are then invited to revise their reference basket by keeping, removing, adding or substituting the products selected in stage 1. The new basket built is called *logo basket*. Finally, participants are asked to fill a questionnaire in stage 3.

Incentives

To avoid hypothetical bias (Carson and Groove 2007; Carlsson 2011), decisions are made incentive compatible. In particular, participants were aware before coming to the lab that they would have the opportunity to buy food products for research purposes. In practice, participants are given at the outset of the experiment €25 as a fixed compensation for participating in the study. At stage 1, they are informed that at the end of the session they would have to buy a significant sub-set of the basket filled during the session: one eighth of the content. The products actually bought correspond to the products available in the laboratory's back office. Participants do not know the products available in the lab until the very end of the session. At stage 2, participants are then informed that only one of the two baskets (reference basket or logo basket) would be randomly chosen for the actual selling at the end of the session. At the end of a session, each participant buys one eighth of the products from their randomly drawn basket. They pay for these products at the prices posted in the catalogue and go home with them. Subjects spent on average less than €5.

Subjects Recruitment and the Organisation of the Sessions

Participant recruitment was done via telephone, Internet and flyers. The experiment was conducted with a representative sample of 364 adults in the greater Grenoble area in France. Participants had to be 18 years old or older, to have at least one child living in the household, and to be a regular food shopper for the entire household. They were aware that the research was public, aimed at fundamental research and that no private corporation was involved in any way in the research, including its objectives and funding. The nutritional aim of the study was not mentioned before the second stage of the experiment. Participants were therefore not told that the French Ministry of Health had funded the research. All sessions were held between 6 September and 9 October 2010. Sessions took place in the experimental laboratory of the Grenoble Institute of Technology. Forty-four sessions were organized, each dedicated to one of the 7 treatments. A session lasted two hours. Each participant was seated, alone, in front of a personal computer screen and paper catalogue.

Data analysis

This study aims at measuring the nutritional impact of different logo formats on food shopping baskets. We first measure the relative distance (in %) between the reference basket and the logo basket for each subject, and thus the changes, *ceteris paribus*, induced by the logo (*within subject* method). On this basis, one can measure the relative effectiveness of the seven logos by comparing the extent of changes between logo formats (*between subject* method). This is possible because each treatment differs only in respect to the logo. We consider not only average distances but also individual dispersion. In particular, individuals who improve, do not alter and reduce the nutritional quality (perverse effect) of their baskets are distinguished.

In order to estimate the nutritional quality of a shopping basket, the LIM score proposed by Darmon et al. (2009) is used. The LIM score is a standard index used by nutritionists to estimate the mean percentage of the maximal recommended values for free sugar, salt and SFA. In other words, the LIM score averages the content per 100g of free sugar, salt and SFA weighted by the nutrients' daily maximal recommended values. It is calculated as follows: LIM=100 × {(free sugar)/50 + (Sodium/3153) + (SFA/22)} / 3. Our change indicator between stage 1 and stage 2 is the distance in percentage between the LIM based on the reference basket and the LIM based on the logo basket.

Non-parametric tests to define statistically significant results are used. With matched data (distance between reference basket and logo basket per individual) we use the Wilcoxon matched-pairs signed-ranks (WSR). For unmatched data (distance between reference basket and logo basket per logo, option or subjects' characteristics), we use the Mann-Whitney test (MW). Finally, we use the Fisher Exact test (FE) when proportions are compared.

Results

A uniform, standardized front-of-pack logo exhaustively implemented on food supply induces a significant improvement of the nutritional quality of shopper's household baskets, irrespective of the logo format. Logo format matters as behavioural responses strongly depend on its components. GDA is neither the best nor the worst possible format; though it is the best among nutrient formats. The simplest is the best: efficient logos use an aggregated ranking. Intra category substitutions occur more frequently with a category subset format, but they are less effective in improving nutritional quality. Category and overall formats eventually appear as equally efficient. On average traffic light display induces responses with much better

nutritional effect than a green display. Unfortunately, traffic light formats also have unintended consequences on a significant share of participants.

Result 1. Overall the implementation of a logo induces significant nutritional improvement

Logos matter, as they induce actual responses in shopping behaviour and changes improve the average nutritional quality of household baskets as measured in the lab. A uniform, standardized front-of-pack logo, exhaustively implemented on food supply induces a significant improvement of the nutritional quality of a shopper's household basket, whatever the format of the logo. On average, when pooling the seven treatments data, results exhibit a statistically significant (WSR, *p*-value=0.000) LIM decrease of -8.7%, *i.e.* a nutritional improvement. 68.1% of the participants improved the nutritional quality of their basket, by a median reduction of -10.7%. 12.1% of the participants remain unchanged and 19.8% reduced the nutritional quality of their basket when a logo is implemented (median LIM decrease by -3.6%).

Result 2. The 7 formats tested reveal highly contrasting efficiencies

Logo format does matter. Behavioural responses strongly depend on the format implemented, and some are much more efficient than others. In descending order the mean changes (LIM ratio) are: POT (-14.6%), PCT (-13.8%), POG (-10.8%), GDA (-10.6%), PCG (-9.7%), NOG (-5.2%) and NCG (-4.4%). The two nutrient logo formats are the least effective; the two traffic light product logo formats are the most effective (Top graph in figure 1 and table 2).

Result 3. GDA is not the most effective format but the best among the nutrient formats

GDA is neither the best nor the worst possible format. Its performances are average. However it is the best among nutrient formats. GDA average performance (LIM indicator) ranks fourth among the seven formats with an average decrease of -10.6%. With it 76.9% of participants

improved their nutritional performance while 15.5% deteriorated it. GDA is better than the other two nutrient logos (NCG and NOG): its overall impact is significantly higher (-10.6% against -4.8%) and perverse effects affect significantly fewer participants (15.4% against 25.5%, F, *p*-value=0.094).

Result 4. Product option is more efficient than nutrient option

In logo terms, the simplest is the best: an efficient logo is simple. When the entire product is qualified with an aggregated index, a format is twice as efficient as when each nutrient is qualified separately: the overall impact on the LIM is -10.2% on average for the 'product' pool (POG plus PCG) and -4.8% for the 'nutrient' pool (NOG plus NCG) (MW, *p*-value=0.000). Option P induces the largest improvement (LIM decrease over -20%) for a greater proportion of participants: 81.1% against 53.1% for the option N (FE, *p*-value=0.000). With option N more participants do not change their basket – neutral effect – (21.4% *vs.* 6.3%, FE *p*-value=0.001) and even more participants decrease the nutritional quality of their baskets (25.5% *vs.* 6.3%, FE *p*-value=0.011) (Second graph in figure 1).

Result 5. Reference set options induce balanced effects and heterogeneous behaviours

The difference between the two options C and O (category and overall) is not significant. On average O induces a -7.5% LIM decrease and C a -6.8%. Although the overall impacts are similar, it stems from different behaviours: Option C induces more substitutions than option O, but most substitutions under the C option are intra-category substitutions and most substitutions under the O option are inter-category substitutions. On average subjects substitute 2.9 products with C option and only 1.7 substitutions with O. However 78% of substitutions with C – compared with 48% with O – take place within the same categories and such category substitutions generate weaker nutritional gains per unit of substitution (Table 3a and 4). Both options generate perverse effects but for different reasons. With O, participants

with perverse effects are clearly non-compliant: they decrease the share of green products (Table 3b). With C participants with perverse effects behave with compliancy: the frequency of green products increases. However, such behaviour would have led to a decrease in green products with an O option as participants have a tendency to replace bad products from good categories with good products from bad categories (Third graph in figure 1).

Result 6. Traffic light option is more efficient but generates more perverse effects

On average traffic-light option T is nutritionally more effective than option green G. But T induces important perverse effects on a significant number of participants. These unintended consequences are not entirely due to a compliance refusal effect. The difference between the two options is 40%: T leads to an average LIM decrease of -14.2% against -10.2% for the G. Is the result significant? According to average yes, but not according to rank (t-test, *p*-value=0.070; Mann Whitney, *p*-value=0.424). On the one hand subjects with favourable effects T is more efficient (LIM decreases by -21.5% against -13.7% with MW, *p*-value=0.080). On the other hand, unintended effects are greater with T (LIM increases by +6.0% against +4.9% (MW, *p*-value=0.084). T generates more extreme effects both ways, more favourable effects but also more perverse effects (bottom graph in figure 1).

Limits and Discussion

As our protocol includes a wide set of data and variables, many choices we have made may be questioned and criticized. (a) By focusing exclusively on shopping behaviour, we excluded consumption, the eating process, which of course is the very determinant of a diet. (b) The protocol architecture chosen, ‘within subjects’ and ‘before and after’, allows for direct individual measurements of behavioural responses. But this protocol induces a high logo

saliency. Results have to be taken as an upper limit behavioural response. (c) Many other formats were possible, we stick to the one 'in the air' of public authorities. (d) Our 273-product selection was small and certainly *ad hoc*, though selected very carefully, it limits the generalisation of our results. (e) Our categories are also certainly *ad hoc* though sorted according to a standard (OQALI). (f) Incentives may be judged insufficient, though meticulously calibrated and welcomed by participants. It was practically difficult to do more without counter-productive effects or rejection (g) Nutrition back-of-pack panels were not available to participants. (h) We did not test a stage 2 with no logo in order to check the neutrality of the protocol architecture. A study is in progress to verify this. (i) Logos should be efficient in contexts of hurried and unfocused real-life shopping. Further research could explore the effects of time constraints or additional tasks introduced while shopping, on the logo formats' relative efficiencies. (j) Calorie density is a simple and easy to understand variable[11]. It is strongly correlated with our LIM indicator. It is a good candidate for alternative formats. (k) Serving may be a better denominator for a nutritional ratio than 100gr but French standards are still in progress. (l) Finally, we exclusively focused on the demand side response to logo. Of course, the strategies of firms will react to such logos. This is out of reach of the present paper. Despite these limitations, we have proposed here a novel experimental protocol capable of measuring, without noise and without hypothetical bias, the effect of global labelling policies that would be very difficult to evaluate in a real-world marketplace. We have obtained significant results that inform the future of the debate on front-of-pack logo formats. First, because logos matter and because the nutritional impact of their implementation is positive, one should certainly continue to support the induction of some sort of front-of-pack logo. Second, as changes in behaviour and induced changes in nutritional quality of food strongly depend on the constituents of the logo, logo formats matter. GDA must not be taken for granted and it is legitimate to challenge this format.

Conclusion

This study sets out to answer two questions. (a) Could a promising, well-designed choice-logo format challenge GDA? (b) Does shopper behaviour accurately respond to the specific stimulus of various formats, and what are the induced effects of these responses on shoppers' basket nutritional quality? Original consequential framed field experiment architecture was designed enabling us to answer these two questions. Based on a 273-item representative selection of the present French food mass supply, a small self-service e-shop was created in the lab. Participants – mothers and fathers who are regular household shoppers – had to consecutively fill two alternative baskets for consumption over several days for the family. Participants fill a first shopping basket – the reference basket. Then, with seven treatments, seven different nutrition logo formats were implemented to mimic labelling systems, i.e. uniformly and exhaustively. Subjects then fill a second shopping basket by, if they wish, revising their initial reference basket. Decisions were made incentive compatible: at the end of the experiment, participants had to buy one eighth of one of the two randomly drawn shopping baskets.

Our study revealed significant food shopper's behavioural response to the implementation of front-of-pack nutritional logo. On average, such behavioural response induces a significant improvement of the nutritional quality of household baskets, as measured with standard indicators (LIM, or energy density). This suggests that a properly designed format of a labelling system may contribute to better diets.

Our main purpose was to study comparative behavioural responses and relative nutritional impacts of various logo formats, so as to appraise their relative performances. A significant result is that GDA, despite its complex underlying heuristic, exhibits good performance.

Nonetheless, and this may be our main result, a simple choice-logo format ranking a product as a whole induces more relevant shopper reactions and has a better nutritional impact. Should products be ranked with reference to its product category or rather to the overall product set? A significant result is that behavioural responses to these two options are significantly different. Shoppers react to each format with accurate precision and compliance. Intra-category substitutions generate larger responses. But as category substitutions have a smaller nutritional impact, the global efficiencies of the two options are similar. A Traffic-light display, signalling both best and worst products has, on average, a better effect on nutritional quality than a green display. But traffic-light formats have significant unintended effects on a large number of the participants. These unintended effects are not only due to non-compliant responses.

References

Allais, O., P. Bertail, and V. Nichèle. 2010. The Effects of a Fat Tax on French Households' Purchases: A Nutritional Approach. *American Journal of Agricultural Economics* 92: 228-245

Ariely, D. 2009. *Predictably Irrational, The Hidden Forces That Shape Our Decisions*. Harper Collins.

Ariely, D. 2010. *The Upside of Irrationality: The Unexpected Benefits of Defying Logic at Work and at Home*. Harper Collins.

Bleich, S.N., D. Cutler, C. Murray, and A. Adams. 2008. Why is the developed world obese? *Annual Review of Public Health* 29: 273-95.

Campbell-Arvai, V., J. Arvai, and L. Kalof. 2012. Motivating Sustainable Food Choices: The Role of Nudges, Value Orientation, and Information Provision. *Environment and Behavior*, XX(X): 1-23

Caraher, M. and G. Cowburn. 2005. Taxing food: implications for public health nutrition. *Public Health Nutrition* 8: 1242-1249.

Carlsson, F. 2011. Non-Market Valuation: Stated Preference Methods in *The Oxford Handbook of Food Consumption and Policy*, ed. Jayson Lusk, Jutta Roosen and Jason Shogren, 716-746. Oxford University Press.

Carson, R.T., and T. Groove. 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics* 7(1): 181– 210.

Chou, S.-Y., I. Rashad, and M. Grossman. 2005. Fast-Food Restaurant Advertising on Television and Its Influence on Childhood Obesity. Working Paper 11879, National Bureau of Economic Research.

Cutler, D.M., E.L. Glaeser, and J.M. Shapiro. 2003. Why have Americans become more obese? *Journal of Economic Perspectives*, 17, 93–118.

Darmon, N., F. Vieux, M. Maillot, J.-L.Volatier, and A. Martin. 2009. Nutrient profiles discriminate between foods according to their contribution to nutritionally adequate diets: a validation study using linear programming and the SAIN, LIM system. *American Journal of Clinical Nutrition* 89: 1227–1236.

DellaVigna, S. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2): 315–372.

Drichoutis, A., R. Nayga, and P. Lazaridis. 2011. Nutritional Labeling, in *The Oxford Handbook of Food Consumption and Policy*, ed. Jayson Lusk, Jutta Roosen and Jason Shogren, 716-746. Oxford University Press.

Drichoutis, A., P. Lazaridis, and R. Nayga. 2006. Consumers' Use of Nutritional Labels: A Review of Research Studies and Issues. *Academy of Marketing Science Review* 9: 1-22.

Etilé, F., 2011. Food Consumption and Health, *The Oxford Handbook of Food Consumption and Policy*, ed. Jayson Lusk, Jutta Roosen and Jason Shogren, 716-746. Oxford University Press.

Grunert, K.G., and J.M. Wills. 2007. A review of European research on consumer response to nutrition information on food labels. *Journal of Public Health* 15: 385–399.

Harrison, G.W., and J.A. List. 2004. Field Experiments. *Journal of Economic Literature* 42: 1009–1055.

Kelman, C. 1958. Compliance, Identification, and Internalization: Three Processes of Attitude Change. *The Journal of Conflict Resolution* 2 (1) : 51-60.

Huang, K.S. 1996. Nutrient Elasticities in a Complete Food Demand System. *American Journal of Agricultural Economics* 78: 21-29.

Just, D. 2006. Behavioral Economics, Food Assistance, and Obesity. *Agricultural and Resource Economics Review* 35(2): 209-220.

Just, D., and C. Payne. 2009. Obesity: Can Behavioral Economics Help? *Annals of Behavioral Medecine* 38(1): 47-55.

Loureiro, M.L., and R.M. Nayga. 2005. International dimensions of obesity and overweight related problems: An economic perspective. *American Journal of Agricultural Economics* 87 (1): 1147–1153.

Swinnen, J.F.M., and T. Vandemoortele. 2011. The Political Economy of Food Standards, in *The Oxford Handbook of Food Consumption and Policy*, ed. Jayson Lusk, Jutta Roosen and Jason Shogren, 716-746, Oxford University Press.

Thaler, R., and C. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

Wansink, B. 2004. Environmental Factors that Increase the Food Intake and Consumption Volume of Unknowing Consumers. *Annual Review of Nutrition* 24: 455-479.

Wansink, B., and J. Sobal. 2007. Kitchenscapes, Tablescapes, Platescapes, and Foodscapes. Influences of Microscale Built Environments on Food Intake. *Environment &Behavior* 39(1): 124-142.

WHO (World Health Organization). 2003. Diet, Nutrition and the Prevention of Chronic Diseases. RePOGt of a WHO consultation: WHO Technical RePOGt Series No. 916, Geneva.

Table 1: Three criteria, six options and 7 logo formats

		Options	
Criteria	**Criterion 1. Units ranked**	Product (**P**)	Nutrient (**N**)
	Criterion 2. Reference sets	Category (**C**)	Overall (**O**)
	Criterion 3. Scopes of colours	Green (**G**)	Traffic-light (**T**)
The 7 Logo formats			
1-GDA	***Guideline Daily Amounts*** displays the percentage of daily-recommended intake values per serving for each of our three nutrients (salt, free sugar and saturated fat).		
2-NCG	***Nutrient Category Green*** displays no symbol 1, 2 or 3 green dots. An additional green path is displayed (with the explicit nutrient name) whenever the product is among the best third of its category subset of products; the ranking appears three times, once for salt, free sugar and saturated fat respectively. Otherwise the logo is left blank.		
3-NOG	***Nutrient Overall Green*** displays no symbol, 1, 2 or 3 green dots. An additional green dot is displayed (with the explicit nutrient name) whenever the product is among the best third of the overall set of products in terms of its contents in salt, free sugar and saturated fat. Otherwise the logo is left blank.		
4-PCG	***Product Category Green*** displays no symbol or 1 dot. A green patch is found when the product is in the best third of its category subset of products concerning its average nutritional quality. Otherwise the logo is left blank.		
5-POG	***Product Overall Green*** displays no symbol or 1 dot. A green patch is found when the product is in the best third of the overall set of products in terms of its average nutritional quality. Otherwise the logo is left blank.		
6-PCT	***Product Category Traffic-light*** includes 1 dot that may be green (same as 4-PCG), or red when the product is in the inferior third within its category products subset. No dot is used when the product is in the average one-third group.		
7-POT	***Product Overall Traffic-light*** includes 1 dot that may be green (same as 5-POG, or red when the product is in the inferior third within its Category products subset. No dot is used when the product is in the average one-third group.		

Table 2: Logos' global and individual impact per treatment

Logos	Average LIM decrease, as % of the reference LIM (standard deviation)	Individual change in LIM from reference to logo basket Percentage of participants in each category			
		-20%< Δ < 0% Improvement	*Δ < -20%* *Large improvement*	Δ = 0% Unchanged	Δ > 0% Degradation
1-GDA	-10.6% (13.5)*	76.9%	*19.2%*	7.7%	15.4%
2-NCG	-4.4% (9.2)*	61.4%	*5.7%*	15.7%	22.9%
3-NOG	-5.2% (15.9)*	45.3%	*12.0%*	26.7%	28.0%
4-PCG	-9.7% (13.6)*	81.4%	*13.6%*	5.1%	13.6%
5-POG	-10.8% (16.0)*	80.8%	*23.1%*	7.7%	11.5%
6-PCT	-13.8% (17.5)*	72.4%	*34.5%*	3.4%	24.1%
7-POT	-14.6% (20.7)*	74.0%	*33.3%*	3.7%	22.2%
All logos	-8.7% (15.0)*	68.1%	*17.0%*	12.1%	19,8%

* Wilcoxon significant at 1%. In red: % of participants for whom the presence of a logo leads to a deterioration of the basket nutritional quality.

Tables 3:

a. Average number of items per basket in the reference basket and in the logo basket according to the seven treatments

Reference basket	GDA	PCG	POG	NCG	NOG	PCT	POT
Average number of items per **basket** in the reference **basket**	22.3	20.5	23.6	22.2	20.7	19.9	20.4
Logo basket	GDA	PCG	POG	NCG	NOG	PCT	POT
Average number of items per **basket** in the logo **basket**	20.4	20.1	22.2	21.9	19.8	18.4	19.0
Average number of items per **basket** kept from the reference **basket**	18.5	16.9	20.3	19.3	18.6	15.5	17.1
Average number of items per **basket** substituted within the same category	1.4	2.6	0.7	1.9	0.9	2.3	1.0
Average number of items per **basket** substituted across different categories	0.6	0.7	1.2	0.7	0.6	0.6	0.9

b. Changes in the number of products from reference basket to logo basket for the treatments POG, POT, PFG and PFT (as percentages of the number of products in the reference basket)

Behavioral Response	*Changes according to colored products with overall reference*				*Changes according to colored products with category reference*			
	Participants improving the nutritional quality of their basket		*Participants lowering the nutritional quality of their basket*		*Participants improving the nutritional quality of their basket*		*Participants lowering the nutritional quality of their basket*	
Logos	*Green*	*Red*	*Green*	*Red*	*Green*	*Red*	*Green*	*Red*
POR	**+24%**	-18%	**-8%**	+3%	+7%	-16%	-6%	-3%
POE	**+5%**	**-47%**	**-6%**	**-6%**	+6%	-25%	+18%	-4%
PFR	+9%	-15%	-4%	0%	**+62%**	-32%	**+37%**	-3%
PFE	-2%	-20%	-2%	+3%	**+37%**	**-56%**	**+39%**	**-36%**

Bold type: changes directly suggested by the logo applied in the treatment.

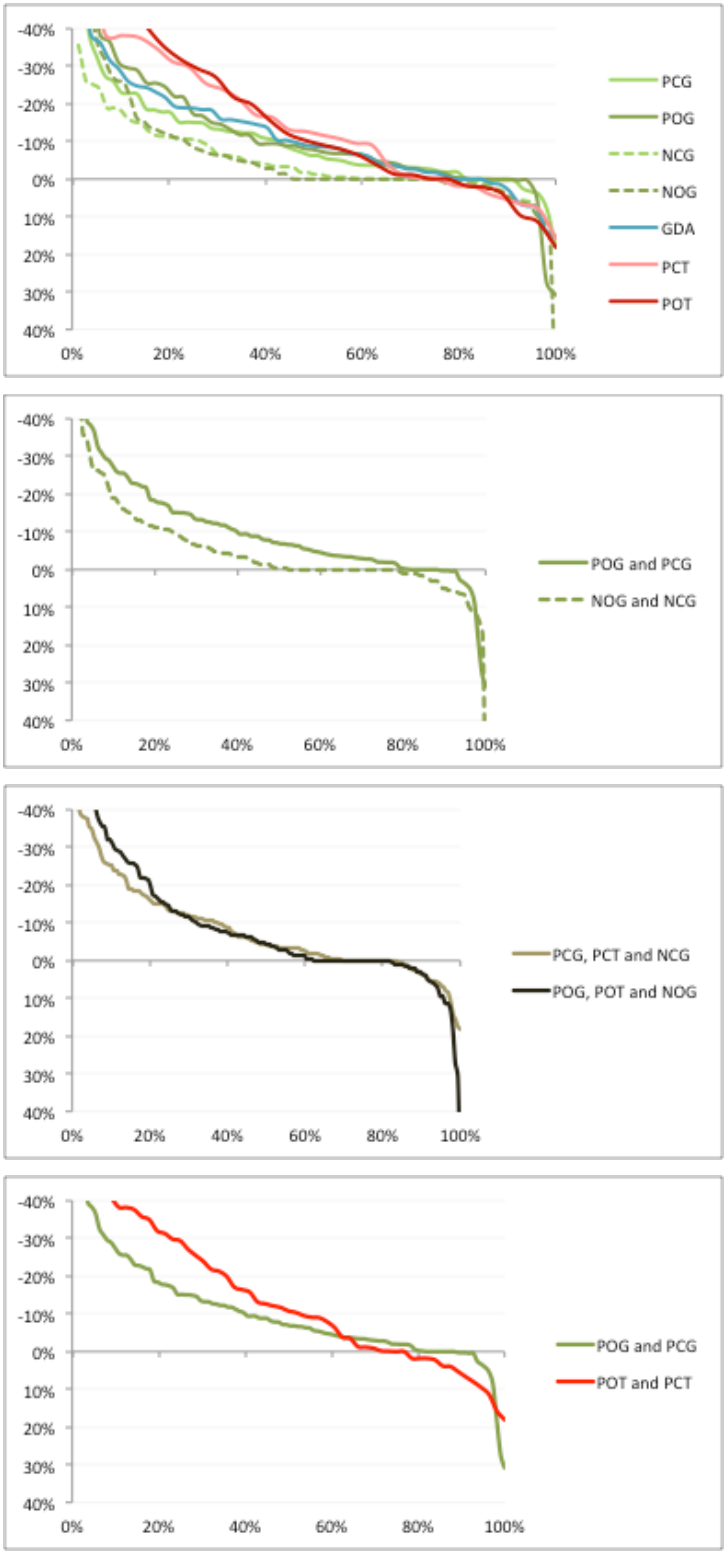
Red type: changes in the direction opposite to that explicitly suggested by the logo applied in the treatment.

Bold frame: logo analyzed through its structure (*e.g.* overall logo formats analyzed through overall structure).

Table 4: Average Nutritional Quality per patch color between the family option and overall option.

		Energy (kCal per 100g)	SA (g per 100g)	AGS (g per 100g)	NA (mg per 100g)	LIM
PFR and PFE	No Patch	212.62	2.96	1.96	116.91	6.17
	Green	176.02	2.60	1.11	115.36	4.64
	Red	245.93	4.06	2.82	194.20	9.04
POR and POE	No Patch	158.11	1.23	0.72	152.10	3.52
	Green	105.81	0.34	0.21	50.23	1.08
	Red	349.75	7.49	4.61	205.96	14.16

Figure 1: Distribution Function of the LIM's variation between the reference basket and the logo basket (in percentage points) per treatment and per option



[1] http://www.gmaonline.org/issues-policy/health-nutrition/facts-up-front-front-of-pack-labeling-initiative/

[2] Breaking down nutrients may also give more transparency and thus appear as more credible.

[3] Data generated by OQALI (http://www.oqali.fr/oqali_eng) suggest (but this might require more analysis) that inter-categories substitutions have better potential, on average.

[4] The Steering Committee of the study included scientists, government officials, and representatives from consumer associations. It validated: (i) the logo formats selection, (ii) the catalogue including products and categories, (iii) the architecture of the experiment, (iv) the criteria used to recruit the subjects.

[5] WHO (Global Strategy on Diet, Physical Activity and Health 2004); USDA (U.S. Department of Health and Human Services, 2010. *The Surgeon General's Vision for a Healthy and Fit Nation.* Rockville, MD: U.S. Department of Health and Human Services, Office of the Surgeon General) and PNNS (http://www.mangerbouger.fr/bien-manger/les-9-reperes/)

[6] Note that the fresh fruits category and the fresh vegetable category, though present in our e-shop, were not labeled in any format treatment.

[7] Condiments and cooking fat are excluded.

[8] The 35 product categories (and number of items in each category) are: Meat (9); Cooked meat (9); Sausages (6); Fish and Seafood (9); Fresh Vegetables (9); Fresh Fruit (9) Canned Vegetables (9); Cooked Potatoes (9), Pasta and Rice (6), Cooked Vegetables (6), Ready-Made Meals with Meat (9); Ready-Made Meals with Cheese (6), Ready-Made Meals with Fish (6) Pizzas (6); Pies and Quiches (6) Sandwiches (6) Snacks (9), Salads (6) Soups (6), Milk (9), Dairy Products (9), Cream (6) Pudding (6); Flavoured Milk Dessert (6), Dry Cheese (9),

Bread and Buns (9), Pastries and Cakes (6) Biscuits (9) Chocolate Bars (6), Breakfast Cereals (9) Spreads (6) ; Chocolate (9); Fruit Desserts (9); Juices and Nectars (12), Sodas and Syrups (12).

[9] OQALI is a French public agency whose mission is to observe food quality, and in particular nutritional quality (http://www.oqali.fr/oqali_eng).

[10] The front-of-pack pictures of each of our selected 273 items, the logos, the graphical user interface, and the store paper catalogues were designed and produced in our lab, specifically for the purpose of this experiment.

[11] European Food Information Council, 2006. An Energy-based Approach to Nutrition Information on Food Labels, *European Food Information Council Forum*, 3, July 2006.