Food labeling for human beings.


Paolo Crosetto, Laurent Muller, Bernard Ruffieux
INRA, GAEL, Université de Grenoble, France

This paper contributes to the debate on front-of-pack nutritional labeling for food. We implement an incentivized laboratory experiment to assess the relative performance of Guideline Daily Amount (GDA) and Traffic Light (TL) labeling schemes in assisting consumers to build a healthy daily diet.

GDA provides detailed nutritional information per serving, as % of an officially recommended daily consumption basket. TL displays color-coded coarser information about the amounts of key nutrients. From a behavioral point of view, GDA and TL sit at two ends of a continuum. GDA gives detailed information, but requires high computational and memory skills to be effective in shaping consumer behavior. TL gives coarser information, but requires lower cognitive skills. GDA is designed for careful, skilled, unconstrained decision makers, and would be perfect for homo oeconomicus. Behaviorally, though, it has been observed that too much information can lead to bad choices, especially if the information is multidimensional (Greifeneder et al., 2010); TL could fare better for cognitively constrained consumers.

Previous studies (Kelly et al., 2009; Moeser et al., 2010; Grunert et al., 2010) have compared the performance of GDA vs. TL (see, for a review, Grunert and Wills, 2007). Most of the existing literature, though, is based on simple tasks consisting in ranking two products according to their perceived healthiness. The problem with these approaches is their focus on a limited number of products, and their non-incentivized nature. Consumers in the shops face a much more complicated task on at least two levels. First, consumers must rank dozens of products that differ in several characteristics and are competing for their limited attention at the same time, while shopping for several items in order to build the food stock for the next days or week. Second, consumers must keep a running memory of what they have bought, in order to fully take into account the overall nutritional balance of their shopping. Moreover, a thorough survey of the literature reveals that the question asked determines the relative performance of the labeling schemes. When the subjects are asked to rank the products’ relative healthiness, or to classify them into three-way scales as healthy/medium/unhealthy, TL wins. When instead the subjects are asked absolute levels, i.e. to evaluate how much of a nutrient is present in each product, GDA wins. This is not surprising: labels perform best when the question asked is the one they have been designed to answer.

But the real goal of nutritional labels is to help consumers to build healthy diets. In our experiment we ask the subjects to do exactly that. Our subjects act as hired nutritionists of a refectory. They must compose a daily diet, choosing from a predetermined set of products, and are paid in cash only if the diet they built satisfies a set of nutritional goals. These can be 1- (kcal), 4- (kcal, fat, sugar and salt) or 7-dimensional (kcal, fat, sugar, salt, fiber, vitamin C and calcium). To guide the subjects in their choices, nutritional labels are provided. We implement three treatments: GDA, TL, and a mixed GDA+TL label.

We implement two different experiments with the same overall structure.

In a first experiment, we want to test the extent to which GDA works when employed by real subjects in a complex task. We hence give the best chances to GDA: subjects have unlimited time to complete their task and are endowed with paper and pencil to perform calculations. We recruit two different samples of subjects: engineering students, and a representative sample of the general population, both from Grenoble, France. We expose them to 15 different daily diet tasks, and measure their performance and time spent.
Results of both the student and general population samples show that GDA performs dramatically better than TL, on both 4- and 7-dimensional tasks. Time spent on each task is high (3 minutes on average). TL leads to slightly faster decisions than GDA. Both samples take long time and perform extensive computations, the main difference being that engineering students are not surprisingly better at constrained optimization than the average person.

A second experiment is meant to test the robustness of these findings, and their external validity. We introduce two changes: subjects are no more given paper and pencil and they face a limited time of 2 or 5 minutes, depending on the treatment. We hence depart from experiment one in the direction of a more realistic setting. We recruit subjects from the population at large. This second experiment is underway. We hypothesize that GDA will perform worse, as subjects are forced by the absence of tools and by the time limit to switch to heuristics and intuition. If this effect is large enough TL might outperform GDA. The exact level of complexity at which TL will overcome GDA, if at all, will guide our final assessment of the two labeling schemes. The interaction between the labels and the cognitive skills of our participants will allow us to assess whether GDA is good also for *homo sapiens*, or if, instead, the slightly paternalistic but simplified approach of TL must be preferred.

References


