Regulating Online Choice Architectures Using Induced-Value Experiments

Damien Mayaux*

May 21, 2025

Summary

This note proposes an experimental procedure to measure the effect of online interface elements - such as marketing cues, featured product box, dark patterns - on consumer welfare. This approach does not require market data or algorithmic disclosure. It quantifies in monetary terms the potential harm or benefit that a marketing practice can cause to a reasonably informed consumer. It can also be used pre-emptively to test whether the presence of a given interface element or a variation of its visual appearance would benefit consumers.

The methodology is based on an artifical choice task performed in an environment that is visually similar to the original interface, in which participants are paid based on their ability to choose according to a prescribed rule. The prescribed rule serves as a reference for how participants try to choose in the experiment. When they mistakenly deviate from the rule, they incur a monetary loss that can be interpreted as a harm. By comparing the monetary loss of participants under different interfaces, the experimenter can quantify the harm caused by a specific interface element.

This procedure has been tested in two experiments (N=140 and N=276 participants) on marketing cues found on the website of a large food retailer. Results indicate that visual cues can significantly affect consumer welfare. Mere variations in color had no statistically detectable effect between groups of 45 participants, suggesting that larger sample sizes are required to assess subtle visual changes. Replications with larger samples size and on other marketing practices are needed to establish the robustness of the results obtained by this procedure.

Figure 1: Examples of visual marketing cues

Context

E-commerce websites can rely on A/B testing to adjust their interface and recommendation algorithms so as to maximize their business metrics - such as click probability or conversion rate. There is no direct equivalent that regulators or consumer organizations could use to serve consumers' interests.

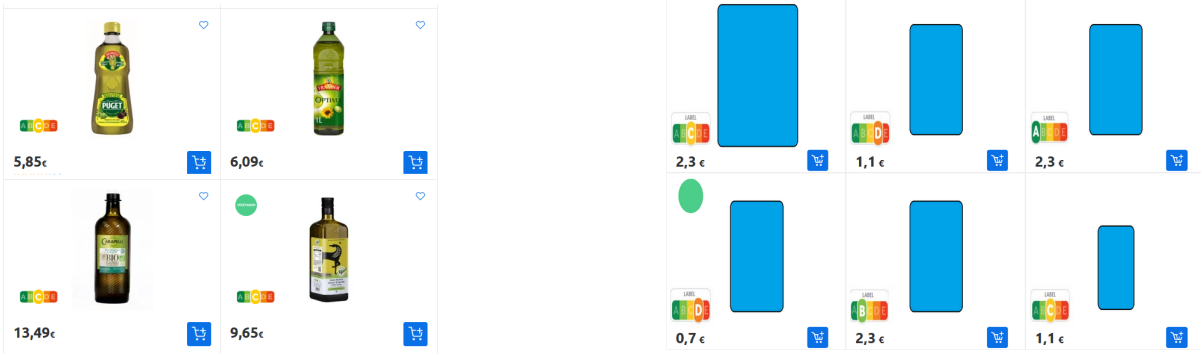
Measuring the effect of a personalized interface on consumer welfare using real-world data is a difficult, lengthy and costly task. It requires some understanding of what the algorithm does and - more importantly - of what consumers want. This is already challenging for e-commerce websites, which somehow understand their algorithms and have a lot of data about their consumers. It is virtually impossible for regulators or consumer organizations, which need to retro-engineer algorithms, ask the website to access the relevant code and data, and acquire complementary market data to learn about consumer preferences. This may work when preferences are simple and the interface clearly biased - e.g. RGPD cookie banners or subscription cancellation forms. However, it does not apply to most purchase decisions on e-commerce websites.

Still, quantifying the potential harm or benefit of an interface on consumers is extremely useful. Consumer protection litigations often require an assessment of the damage to a reasonably informed consumer. Regulators could detect early on marketing practices that may harm consumers and propose remedies. Benevolent websites could introduce estimates of consumer welfare in their business metrics. If a rough estimate of the welfare effects of an interface could be obtained immediately and at a low cost, the regulation of choice architectures would go much faster and become more pro-active, if not preventive, ensuring a better enforcement of the DSA and saving on downstream investigation and litigation costs. The experimental procedure presented in this note tries to rise to this challenge.

*Paris School of Economics, damien.mayaux@psemail.eu

Methodology

An artificial product choice task with induced preferences. The procedure relies on an artificial product choice task obtained in three steps. (1) Create an interface visually similar to the original website, but without reference to actual products or characteristics - see Figure 2. The choice task should mimick the cognitive requirements of actual purchase decisions. (2) Define how participants are paid depending on their product choice. For instance, give a monetary value to each possible characteristic and define the value of the product as the sum of the value of its characteristics. This payment rule will induce the preferences of the participants in the experiment and serve as a monetary scale of consumer welfare. (3) Decide on the number of choices that participants will have to make in the experiment and the difficulty of these choices. This should reflect the conditions under which users encounter the interface during a single shopping session on the original website.



(a) Original website interface

(b) Experimental interface

The experimental interface is deprived of meaningful content or reference to actual product characteristics, except the price. The visual layout of the page and the interface element under study - here, a green circular marketing cue - are preserved as much as possible. Participants have to choose one of the six "products" in blue. They are paid based on the characteristics of the product. For instance, a large rectangle is worth 3€, a label C is worth 1.5€ and the price has a negative value, hence choosing top left product in the example earns 3+1.5-2.3 = 2.2€.

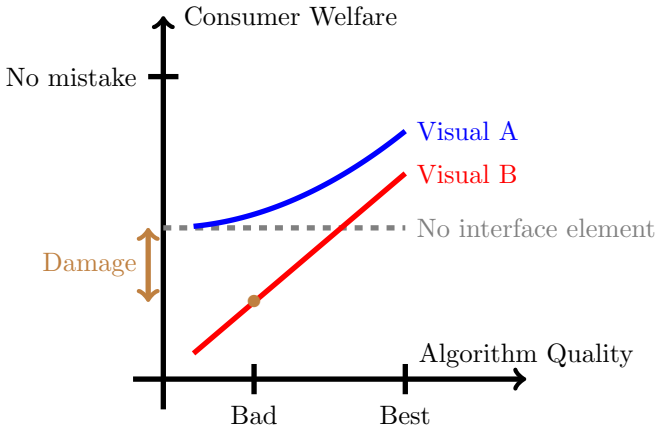
Figure 2: From the original website to the artifical choice task

In this artificial product task, participants know the payment rule and can therefore be seen as reasonably informed consumers with preferences measured in monetary terms. They may make mistakes, in the sense that they do not choose the product that pays the most, and such mistakes correspond to a monetary loss. Now, the objective is to measure how the interface affects this monetary loss, which measures the damage.

Manipulating the interface. Interface elements promoting a product can be described by their visual appearance and their algorithm. Here, "algorithm" means the rule that determines which product is promoted. To explore the set of possible algorithms, it is simpler to describe them in terms of the algorithm quality, that is to say how good the promoted product is with respect to consumer preferences. The best possible algorithm quality consists in always promoting the product that the consumer values the most.

The experimenter can introduce variations in both the visual appearance and the algorithm quality. It is also possible to study choices in absence of the interface element under study. By comparing these different conditions, one can learn how helpful or harmful an interface element can potentially be to consumers. The monetary payoff in the experiment quantifies consumer welfare in each condition.

The outcome of the experimental procedure is a "consumer welfare map" - see Figure 3. The next section presents some consumer welfare maps produced in two experiments and discusses the policy implications.



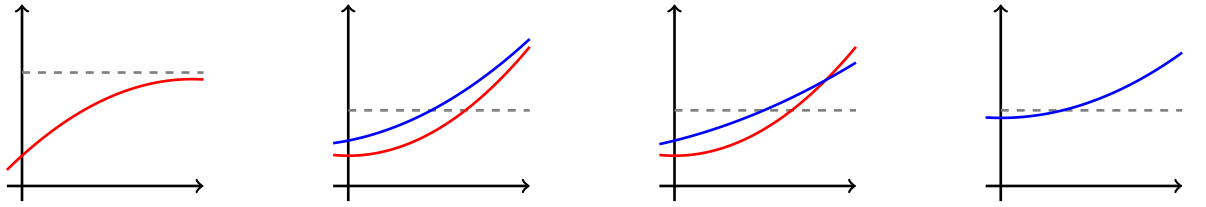
Consumer welfare is measured as the difference in monetary payoff between the chosen product and the best available. The case of "No mistake" gives an upper bound. THe "Best" algorithm gives an upper bound in algorithm quality. The dashed gray line represents consumer welfare in absence of the interface element. The damage corresponds to the difference in consumer welfare between two situations, for instance - in brown here - the absence of interface element versus the visual B with a "Bad" algorithm.

Figure 3: "Welfare map" of an interface element

Results and policy implications

How to read a "welfare map" to address regulatory questions? Some configurations have immediate policy implications. Figure 4a is such that the presence of the interface element is always detrimental to consumers, no matter the algorithm quality. Figure 4b is such that one visual is always better than the other. In both cases, a conclusion about potential interface improvements can be reached without any knowledge about actual consumer preferences or the algorithm used by the website.

Some other configurations require a more careful analysis. Figure 4c is such that a visual is almost always better than the other, except when the algorithm quality is almost perfect. Therefore, the latter visual may benefit consumers only if the website has the will and ability to use an algorithm that aligns well to the individual preferences of each consumer. Figure 4d is such that the interface element can greatly help consumers but barely harm them. The regulator may be willing to take the risk to allow it.

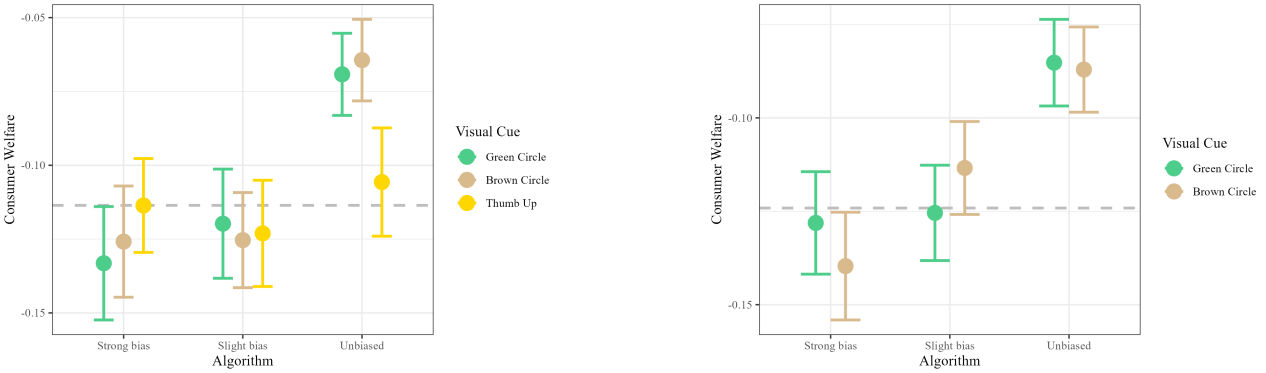


(a) Harmful element (b) Better visual (c) Often better visual (d) Harmless element

Figure 4: Some interesting configurations for a "consumer welfare map"

Results from two experiments on marketing cues. Figure 5 presents the "welfare maps" obtained in two experiments. The experiments investigated the effect of the marketing cues from Figure 1 when encountered repeatedly during a single shopping session. The first experiment suggests that, (1) the thumb up visual has no significant effect on consumer welfare, no matter the algorithm, (2) the green circle could help or harm consumer depending on the algorithm and (3) the brown circle could help consumers but barely harm them. The average welfare loss was 0.11€ per choice in absence of cue and 0.13€ with a bad algorithm, hence a potential damage of 0.02€ per choice.

The second experiment focused on the green and brown circle with a slightly modified design - see the notes of Figure 5 - and more statistical power. The color effects are still not significant and the configuration of the point estimates is different. This suggests that color effects are either sensitive to the experimental design - for instance, how many cues are seen by a single subject - or require larger sample sizes to be detectable.



(a) First experiment (N=140) (b) Second experiment (N=276)

Each subject had to make 10 product choices in presence of each cue and 10 times without cue. The 1st experiment had three cues, hence 40 choices. The second has twwo cues, hence 30 choices. In the first experiment, the algorithm could vary with the cues, whereas in the second, each subject faced the same algorithm for both cues.

Figure 5: Experimental results from two experiments on marketing cues

Discussion. The proposed methodology has a limited implementation cost - running the experiment takes on average 30min, paid on average 7€ per subject, on a dedicated online recruitment platform. It can easily be replicated at scale with other visual elements.. It could also be adapted to determine the welfare map of controversial interface elements - for instance, to determine whether they qualify as dark patterns. However, further methodological work is required to understand the possible sensitivity of the results to the experimental design before this experimental procedure can be used as legal evidence.