HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Date Marks, Valuation, and Food Waste: Three In-Store 'Eggsperiments'

Alessio d'Amato

Timo Goeschl

Luisa Lorè

Mariangela Zoli

AWI DISCUSSION PAPER SERIES NO. 693
November 2020

Date Marks, Valuation, and Food Waste: Three In-Store 'Eggsperiments'*

Alessio d'Amato[1,2], Timo Goeschl[3,4,5], Luisa Lorè[1], and Mariangela Zoli[1,2]

[1]University of Rome "Tor Vergata"
[2]SEEDS, University of Ferrara
[3]Alfred Weber Institute of Economics, Heidelberg University
[4]Heidelberg Center for the Environment, Heidelberg University
[5]ZEW - Leibniz Center for European Economic Research, Mannheim

October 21, 2020

Abstract

We provide causal evidence on how date marking policies influence consumers' valuation of perishable food products through three consecutive research steps. In a preparatory in-store survey ($n = 100$), we identify perishable food items that can be experimentally manipulated to overcome core challenges for causal identification. A modified in-store multiple price list (MPL) experiment ($n = 200$) then tests consumers' valuation of perishable food of varying shelf-life (expiry date) in a two-by-two design that varies date mark type (use-by versus best-before) and information status while preventing free disposal censoring. We find that expiry dates affect consumer valuation, but not differences in date mark type. Educating consumers about date mark meaning turns out to be conducive to discarding potentially unsafe food, but not to preventing food waste. An attentiveness experiment ($n = 160$) tests whether these treatment effects plausibly result from the nature of consumers' knowledge and finds that the existing asymmetry in consumers' understanding of current date marks can explain the evidence from the modified MPL experiment.

Key words: Date marking; food waste; consumer valuation; information-based policies; multiple price list experiment; in-store experiment.

JEL codes: D12, C93, Q53, Q18

*We thank Christina Schmidt for excellent research assistance. We are grateful for helpful comments to audiences at the 8^{th} Annual Conference of the Italian Association of Environmental and Resource Economists in Brescia, February 2020, the 2020 NERI Workshop on Regulatory Economics in Naples, the 2020 online conference of the European Association of Environmental and Resource Economists, and the Grenoble Applied Economics Laboratory as well as to colleagues at the Research Center for Environmental Economics for helpful discussions. Thanks to Bradley Rickard, Norbert Wilson, and Shuay-Tsyr Ho for sharing their experimental instructions. Financial support by the University of Rome "Tor Vergata" under the Mission Sustainability Research Programme and by the Heidelberg Center for the Environment under its Small Grants Scheme are gratefully acknowledged. Correspondence: goeschl@uni-heidelberg.de

1 Introduction

Product labeling is a widely practiced type of information-based policies intended to educate households about important features of the goods they are purchasing and consuming. One application of product labeling that has attracted considerable attention in the current debate around food waste is date marks on food products. Date marks are supposed to play an important role in how food producers communicate actionable information to consumers. An externally visible product feature, their intended role is to inform about key properties of a food item by attaching a label with a calendar date to the product.

Date marking policies for food balance two competing objectives: On resource efficiency grounds, households should retain, and ultimately consume, the food that they have purchased such as to minimize waste. On health and safety grounds, households should discard food that poses an appreciable risk. The solution in most countries has been to mark food products with a calendar date. This date communicates one of two different relationship between the date and food quality, depending on the label type is used. A *best before* (BB) type of date mark is a measure of *food quality*. There, the calendar date shown is intended to communicate the time horizon over which the food item maintains its highest quality, but remains safe to consume within some time after that date.[1] A *use by* (UB) label type, by contrast, is a measure of food safety. It shows a maximum calendar date until which the food item is safe to consume and after which it should no longer be consumed.[2] While both date marks are typically referred to as the *expiry date* and only differ in wording,[3] they anyway pursue very different objectives in informing consumers in an actionable way: The *BB* label communicates a date before which the food product keeps its full consumption value and should typically be retained. The *UB* label communicates a date after which the food product has negative consumption value on food safety grounds and should be discarded.

Among policy-makers, there is a recognition that date-marking policies are a significant contributor to food waste because consumers misinterpret the meaning of date marks. Studies in the European Union (EU) have linked 10% of the 88 million tons of food wasted in the EU directly to date marking, with an estimated cost of €14 billion.[4] Changes in date marking policies and information campaigns to educate consumers about date marks have therefore been high on the agenda, driven by a perception that consumers particularly misunderstand the *BB* date mark (EU-Council, 2016). This current recognition of how consumers relate to date-marking policies has been generated through extensive survey evidence (see review below), which has highlighted many critical issues about the public understanding of date marks and about consumers' statements on food management in the household. To the economist's taste, however, there is a paucity of evidence on how date labeling policies *causally* affect consumers' *valuation* of and

[1] The legal definition in the European Union is that the best-before date is "the date until which the food retains its specific properties" (EU Regulation No. 1169/2011, p. 26)

[2] After that date, "food shall be deemed to be unsafe" (EU Regulation 1169/2011, p. 35).

[3] One could imagine regulators choosing considerably more salient differences between the labels.

[4] `https://ec.europa.eu/food/safety/food_waste/eu_actions/date_marking_en`

choices about food items.

Economists' interest in how consumers value food is natural: Whether food items in possession of the consumer are valued positively vs. negatively maps intuitively onto the retain vs. discard decision. The focus on valuation also responds to empirical challenges for the economist interested in food waste. For obvious reasons, there is a lack of natural experiments that would vary date-marking policies in the field. And observing effects on retain vs. discard decisions requires, at a minimum, researcher access to the content of household waste-bins, which is difficult to obtain for a number of reasons, not least issues of privacy.[5] The lack of causal evidence is the result of additional empirical challenges that researchers encounter in a food labeling context as soon as they move beyond surveys. Ideally, the researcher would be able to manipulate subjects' beliefs and knowledge about and awareness of existing information-based policies. In practice, subjects may hold strong beliefs or possess hard knowledge about the existing policy such that the researcher's experimental treatment fails to induce a different set of beliefs or knowledge. The researcher would also want to manipulate the type of date marks and the calendar dates stamped on the food product. Such external manipulation is important in order to disentangle factors associated with specific foods from the role of the date mark. In practice, unless expertly done, subjects may recognize the mismatch between product type and date label type or detect the manipulation. The researcher will also want to create food management decision contexts that allow the issues of acquiring, keeping, and consuming or discarding food items to establish themselves in a consequential way.[6] Any manipulation, finally, needs to adhere strictly to research ethics and food safety: Consumers may not be exposed to additional food risks as a result of experimental manipulation.

Our experimental approach addresses the empirical challenges to causal inference through a sequence of three steps. First, on subjects' knowledge and beliefs, our experiment is informed by a preparatory survey ($n = 100$) in which subjects disclose their beliefs about date marking of different grocery products and their confidence that their beliefs are accurate. We find a widespread lack of confidence and considerable heterogeneity in the accuracy of shoppers' beliefs about the applicable date mark across grocery items. For some items, shoppers are no better than a coin toss in guessing the correct date mark, have low confidence in their guess, and only a quarter or less of subjects express confidence in their guess. In the subsequent modified in-store multiple price list (MPL) experiment we therefore employ a food item (eggs) for which beliefs on date marking turn out to be demonstrably weak. Second, the product and country setting (eggs, Germany) is conducive to date mark manipulation that is credible and non-detectable: In contrast to other countries, date-marking of individual eggs is unfamiliar to German consumers. The date-mark is instead attached to the outside of egg cartons and can be

[5] One exception may be the case of Sweden where municipalities have been establishing specific food waste collection services. See Ek and Miliute-Plepiene (2018) for a paper that exploits this empirical opportunity.

[6] Typical food items used in experiments, such as chocolate bars (Davis and Millner (2005)), lend themselves to impulse consumption and are typically consumed well within the product lifetime. The issue of discarding food therefore fails to arise.

replaced by the experimenter with acceptable effort, skill, and suitable equipment for producing date mark labels. Third, storage, consumption and discarding are relevant features of consumers valuing eggs: Eggs are not consumed on the spot, but are perishable[7] and easily disposed of. Finally, the setting is conducive to maintaining ethics and food safety: Eggs are, by law, date-marked *best-before*. By forward-dating expiry dates on egg cartons and by affixing a *use-by* date mark, subjects face the same or lower risk in all experimental conditions than they do in the marketplace.

Specifically, we recruit 200 grocery shoppers in a suburban grocery store for the MPL experiment by inviting them individually to participate in an initial survey about food shopping habits in exchange for a combination of monetary (€2) and in-kind (a box of six organic eggs) rewards. After receiving the reward, subjects are randomly assigned to one of four treatment conditions. One condition varies the date mark type (*BB* or *UB*). The other varies the information status of the subjects, either relying on subjects' *native* knowledge or *educating* subjects prior to taking consequential choices. In this two-by-two design, all subjects then make MPL choices over cartons of eggs with expiry dates manipulated by the experimenter. The modification of the standard MPL experiment consists of adding a terminal buy-back mechanism at the end of the experiment. This methodological contribution of the paper allows the researcher to overcome free-disposal censoring in cases in which theory supports both positive and negative valuations for a product. A terminal buy-back mechanism has applications beyond food (waste) to areas such as products that are potentially noisome or dangerous to all or a subgroup of consumers, for example allergenic or repugnant products. The modified MPL experiment delivers for each of the four treatment conditions willingness-to-pay (WTP) data of 50 individuals for eggs with an expiry date seven days after, one day after, on the day, and the day before the subject participates.

The final step in our approach is an experiment that tests whether the treatment effects that we establish in the modified MPL experiment are consistent with the nature of consumers' information about date marks. We recruit 160 shoppers in two treatment conditions that provide incentives to subjects to reveal their native knowledge about date marks and the attention that they pay to the information that policy-makers attach to packaged food products.

Our main results are as follows. First, shoppers' beliefs and knowledge about date marks are suitable candidates for experimental manipulation. Second, experimentally varying the date mark type and consumers' knowledge does not significantly affect consumers' valuation at or around the expiry date, except under specific circumstances: A significant treatment effect of changing the date mark type and of educating consumers only arises when subjects face a product beyond its expiry date. This treatment effect works in the opposite direction, however, to the policy-makers' intentions: Education and a use-by label decrease consumers' valuation of perishable food items, rather than increasing their valuation when a best-before date mark is used. In other words, the likely effect of an education campaign on the correct meaning of date labels is to increase socially "desirable" discarding, namely when a health risk concern is involved,

[7] In addition, eggs' shelf life cannot be extended through freezing.

rather than reduce socially "undesirable" food waste. Third, the cause of this effect can be traced back to consumers' tendency of being better informed, contrary to policy-makers' assumptions, about the meaning of best-before date marks than about use-by date marks.

2 Related literature

Food waste is a topic of considerable importance in Europe and other developed economies (EU-Council, 2016; Hall et al., 2009). Observers take the amount of food waste as indicative of inefficiencies and unnecessary environmental damage, pointing to the amount of water, carbon dioxide, pesticides etc. associated with the production of foodstuffs that never reach their intended destination. Not every instance of food being discarded is necessarily inefficient (Ellison and Lusk, 2018), but the scale of the problem alone justifies more research on the phenomenon. Consistent with the importance accorded to food waste, researchers have begun to develop better estimates of the amount of food waste (Stenmarck et al., 2016), to study the determinants of food waste (Ellison and Lusk, 2018; Daniel, 2016; Secondi et al., 2015; Quested et al., 2013), and to develop policy proposals for reducing food waste (Hebrok and Boks, 2017). Private households are estimated to account for more than 50% of food waste in the EU (Stenmarck et al., 2016), and for 43% of food waste by weight in the US[8]. Yu and Jaenicke (2020) estimate that the average household in the U.S. wastes 31.9% of bought food, corresponding to an annual aggregate consumer-level value of US$240 billion. Well designed date-marking policies could therefore arguably make an important contribution to better align the behavior of consumers with political objectives if households respond to date marking in the intended way.

Consumers report in surveys that they are aware, understand, and act upon date marks on food product (e.g. Eurobarometer, 2015) while researchers consistently find the opposite (Neff et al., 2015, 2019). Evidence suggests that the intended effect of date-marking policies may be limited or can even backfire: Thompson et al. (2018) find in an online survey of 548 Scottish consumers that consumers report no difference in willingness to consume 'expired' food items between use-by and best-before labels. Wilson et al. (2018) show that consumers make unwarranted inferences about food items based on date labels beyond safety and quality, and Roe et al. (2018) find in a lab survey that date marking itself is a plausible cause of food waste.[9] Jointly, these and other results from surveys and studies with consumers cast doubt on the usefulness of date marking perishable food for guiding consumer's purchase, consumption, and waste decisions. European evidence supports this conclusion: Around one quarter of respondents from countries in the EU misinterpreted the *BB* label in a survey as implying that the food is no longer safe after the corresponding date. Such misinterpretation could indeed result in

[8]See, among others, the Refed website: `https://www.refed.com/?sort=economic-value-per-ton`.

[9]In sessions with 88 consumers that could inspect and smell opened milk containers, subjects reported a stronger intention to discard the milk if the container had a date mark compared to containers that did not bear a date mark

"unnecessary" food waste. At the same time, 28% of respondents exhibit an inaccurate understanding of *UB* date marks (Eurobarometer, 2015). If this leads to households consuming potentially unsafe food after its expiration date, there could also be too little food discarded after purchase.

While survey evidence, such as cited above, is informative, it cannot shed light on the causal relationship between date marks and consumers' valuation of food items. There is a small number of experimental studies that provide causal evidence. Collart and Interis (2018) conduct a laboratory experiment ($n = 150$) with general population subjects that can use a $35 endowment to acquire food items of varying shelf life and different perishability, including expired items, in an on-screen choice experiment. Their treatment conditions vary information and expiry dates, but not the label type. Education on date mark meaning does not impact on consumer choice in their experiment, while information relating product choice to food waste and its environmental consequence does. Wilson et al. (2017) conduct a laboratory experiment ($n = 200$) with non-student subjects under four different date labels conditions ("Best by", "Fresh by", "Use by" or "Sell by") and three types of food. An auction is used to elicit both participants' willingness to pay and expected amount of food waste in order to calculate a measure of willingness to waste (WTW), on the basis of the idea that date labels may have an impact on WTW by affecting both WTP and the expected amount of waste. More specifically, the closer the expiry date, the lower the WTP and the larger the expected waste, leading to a larger WTW. The authors suggest that the WTW is larger for labels that imply, in the consumers' perception, a safety concern, namely *UB*, but perform no direct comparison of date labels across subjects. We contribute to these studies in three ways: First, by moving to an in-store setting that gets the experimenter closer to the average shopper; second, by directly manipulating date marks on products, thus establishing a clean treatment effect; and third, by adopting a modified MPL experiment as the elicitation mechanism, overcoming some of the challenges implicit in negatively valued consumption goods.

3 Theoretical considerations and design

The objective of our experimental design is to establish how date mark type and the information level causally affect consumers' WTP for perishable food with different expiry dates, including the possibility of expired food and negative WTP. The guidance provided by the theoretical literature for design and hypothesis development is surprisingly limited.[10] In a much-cited empirical paper, Tsiros and Heilman (2005) provide a review of the literature up to 2005 and conclude that while there is a considerable theoretical literature on supply-side aspects of perishability in the retail sector, the demand side is largely unexplored in applied economic theory. As a result, the authors qualitatively sketch the outlines of a theoretical framework that, fifteen years later, remains little

[10] One exception is an early paper on optimal household inventory management with perishable products (Reinhardt et al., 1973). The focus of the paper's deterministic model is on the implications of product lifetime for shopping frequency, however, rather than the valuation questions explored here.

improved upon.[11] We build on these outlines to inform our experimental design and the formulation of four testable hypotheses on date mark type, information, and expiry date.

Perishability is synonymous with a fixed time horizon over which a good provides positive benefits. Arguing intuitively, Tsiros and Heilman (2005) posit that the instantaneous consumption value (the expected immediate satisfaction derived from consuming the product) is decreasing over time as the good degrades. A slightly richer characterization of the consumer's problem is to consider that the owner of a perishable item faces a trade off between the instantaneous consumption value[12] and the option value, i.e. the value of postponing consumption to a later point. For most food items, both values are directly connected with the expiry date. Distance to the expiry date, i.e. remaining shelf life, weakly increases the perceived quality of the food item and hence the expected consumption value. Likewise, remaining shelf life increases the option value because there is more time left for postponing consumption. Perfectly intuitively, therefore, WTP tends to decrease as a perishable product approaches its expiry date.[13] The evolution of the WTP is captured in its day-to-day changes, or *intertemporal* differentials, and these changes tend to increase in absolute terms as the good approaches (or exceeds) the end of its shelf-life.

Both date mark types target the formation and evolution of consumption and option values, but intend to affect them differently: If effective, a *best-before* mark induces a roughly constant expected consumption value up until the expiry date and a slow decline thereafter, thus giving rise to a positive option value beyond the expiry date and a smooth increase in the intertemporal differentials. A *use-by* date mark, on the other hand, is intended to induce an expected consumption value that is positive up until the expiry date and negative thereafter. If effective, the option value therefore collapses to zero at the expiry date, forcing a jump in the intertemporal WTP differential.

The intended difference between the two date mark types on the formation and evolution of WTP for perishable food constitutes the primary empirical target of the experiment. On theoretical grounds, it is smallest for distant expiry dates and largest at and just beyond the expiry date. Conditions for detecting a treatment effect of date mark type will therefore be favorable when the otherwise identical food items on offer are date-marked close to their expiry date. We also reason that additional education of consumers will render these conditions more favorable yet. This reflects our reading of the prior evidence, cited earlier, that members of the public frequently overlook or misinterpret the date mark.

[11] There is a handful of papers that test psychological theories in the context of perishable food, such as the theory of planned behavior (Siddique, 2012) or value-belief-norm theory (Farr-Wharton et al., 2014).

[12] On contrast to Tsiros and Heilman (2005), one need not require the instantaneous consumption value to decrease strictly with time. It can remain constant for long times, peak at certain times (for example for festive events with traditional food items) or change through time with changes in preferences (e.g. appetite) and external conditions (e.g. weather).

[13] Hypothesis 1 in Tsiros and Heilman (2005) states much the same, without taking the option value explicitly into account, however.

In order to detect an effect of date mark type on WTP, we use a modified multiple price list experiment (Andersen et al., 2006, 2007) in which otherwise identical food items differ, within subjects, by expiry dates and, across subjects, by date mark type. In a MPL experiment, subjects take a sequence of consequential binary decisions with ordered bid prices such that the researcher can identify, with some precision, the switching point between the two choices. One, often unspoken, assumption in standard MPL experiments is that the choice outcomes have non-negative value to the subject. In the context of perishable food, this assumption need not hold, in particular not in the context of food items close to or beyond the expiry date: Subjects may well prefer not to own food that is spoiled or about to spoil. There are two avenues open to the researcher: One is to ensure compliance with an experimental protocol in which food items chosen must eventually and verifiably be consumed by the subject. This is the solution used in the seminal paper by Coursey et al. (1987), in which experimental subjects who win an auction have to hold an unpleasant tasting substance in their mouths for a period of time and in the papers by Shogren et al. (1994) and Hayes et al. (1995) in which subjects have to consume potentially unsafe meat sandwiches in order to collect their participation reward. Even if this approach survived a current ethics review, its feasibility is established only for lab experiments involving student subjects. The other avenue is to augment the design. Here, a simple extension of the MPL into negative price terrain will not suffice due to free disposal censoring. Free disposal means that experimental subjects can always dispose of products by throwing them away at no cost to themselves. This option means that any positive payment in exchange for receiving the unwanted product will be acceptable to subjects. This induces censoring at zero payments: Subjects who prefer to dispose of the good rather than consume it will not be induced to express truthfully their negative WTP even if the MPL includes negative prices. The presence of free disposal censoring can be detected, however, through a simple procedural change: The experimenter offers to buy the item back after the MPL experiment. This terminal buy-back mechanism provides subjects with zero or negative WTP a clear incentive to exchange the unwanted good for a positive reward.

Among the two approaches, our design opts for the terminal buy-back mechanism for three main reasons. One, in an in-store setting with general population subjects, strict protocol compliance is untested and suffers both from questionable ethicality and tenuous enforceability. Two, explaining such a protocol is also likely to have a chilling effect on recruitment of subjects. Three, the buy-back mechanism allows the option value of food products to establish itself: When protocol compliance forces subjects to consume the item on site (to prevent disposal), the option value could not enter WTP and its estimates would understate true WTP for acquiring the product. These three reasons underpin our choice of a modified MPL experiment for eliciting WTP in an in-store experiment for perishable food items.

Food product	Share of correct answers	Median confidence level	Confidence level > 7
Eggs	0.50	5	0.25
Table salt	0.66	6	0.33
Fresh fish	0.64	6	0.31
Sliced bread	0.65	6	0.27
Minced meat	0.66	7	0.42
Fruit juice	0.72	7	0.40
Yogurt	0.90	8	0.51

Table 1: Share of correct answers as to label type, median confidence level and share of strong beliefs, by food item

4 Procedures

4.1 Preparatory survey

Strong beliefs among consumers, whether correct or incorrect, about which foods bear which date mark and high confidence in their beliefs are an obstacle to a successful experimental manipulation of date mark labels. Worse, subjects may be suspicious of labels affixed to the food item that conflict with their highly confident beliefs. In surveys, German subjects claim that they always (51%), often (25%) or at least sometimes (9%) check the expiry label on a product and understand its meaning (Eurobarometer, 2015). This makes German consumers potentially not amenable to experimental manipulation of date mark labels.

To test the viability of an in-store experiments in such a context, we conducted a preparatory survey with general population subjects ($n = 100$) at a grocery store in Germany in order to elicit beliefs, knowledge, and confidence among members of the target population about current date marking practices for a range of foot items.[14] This survey was carried out from Sept. 4^{th} to Sept. 6^{th}, 2019 during morning, afternoon, and evening shopping hours. Subjects were invited to participate in a survey for a small symbolic reward unrelated to their answers. They were handed a tablet computer and were presented, for a selection of seven commonly purchased food products, with a choice of which of the two labels, BB or UB, they believed to be associated with that product. For each choice, subjects were asked to report their confidence in their choice on an 11-point Likert scale from 0 (not at all sure) to 10 (absolutely sure).[15] The survey was not incentivized, reflecting the unsettled debate about incentives in confidence tasks (Lebreton et al., 2018).

Table 1 reports the results of the survey by food category (first column), ranked by

[14] See Appendix B, Figure A1, for a flowchart of the survey.

[15] Other researchers choose a format in which subjects rate their confidence on a scale between 50 % (no confidence=random choice) and 100 % (absolute confidence) ((Lebreton et al., 2018); (Murad et al., 2016)). While appropriate for highly educated student subjects, members of the general population with little familiarity of a quantitative approach are in our opinion likely to find this format challenging.

the share of correct answers (second column). For each category, it also reports the median confidence level among subjects (third column) and the share of subjects with a strong (> 7 on the Likert scale) belief (fourth column). The share of correct answers ranges from no better than flipping a coin (eggs) to 90 % (yogurt). Median confidence across subjects closely tracks the share of correct answers across food items and so does the share of subjects with high confidence in their choice.[16] Among the food categories, eggs stand out as a food item that is a particularly suitable target for an experimental manipulation of date mark labels: Our sample of German shoppers had low median confidence in their guess of the correct date mark type and only a quarter of shoppers held a strong belief in their choice. This is consistent with the fact that only half the subjects picked BB as the correct date label type. These findings provide the basis on which eggs were chosen as the product whose date mark label would be experimentally manipulated in an in-store setting.

4.2 In-store experiment

The in-store experiment was run in a sub-urban grocery store of a medium-sized German city between Sept. 20^{th} and Sept. 21^{st}, 2019 during morning, afternoon, and evening shopping hours. This store belongs to the same chain of supermarkets as the store in which the preparatory survey was carried out, thus attracting the same customer segment. However, it is located 5km away in a different part of the city, reducing overlap between the participant pools of the preliminary survey and the in-store experiment.

To recruit subjects, we invited shoppers walking through the store lobby to participate in a survey in exchange for a reward. Participants were assigned to one of four treatment groups through a randomized assignment protocol.[17] Irrespective of assignment, they completed a survey on grocery shopping behavior and demographics on a tablet computer. The survey was uniform for all treatments and did not refer to date marks or the perishability of food items.[18] As compensation for their time, participants received two rewards at the end of the survey: One reward was €2.00 in coins.[19] The other reward was a carton of six organic eggs date-marked to industry standards with an expiry date always coinciding with the day of the experiment.

The treatment arms separated with the handing-over of the reward: Two of the four treatment groups, labeled *BB*, underwent the MPL experiment handling egg cartons exclusively bearing date marks that read *best before DD.MM.YY*. The other two groups, labeled *UB*, underwent the MPL experiment handling egg cartons exclusively bearing date marks that read *use by DD.MM.YY*. This variation in date mark *type* constitutes the first of two treatment dimensions of the MPL, as summarized in table 2. The

[16] At the individual level, there is only a weak statistical association (Kendall's τ) between the likelihood of correctly answering the question on the label type and the confidence that the subject expresses about her choice

[17] See Appendix B, Figure A2, for a flowchart of the MPL experiment.

[18] See Appendix A for details on the survey questions.

[19] Consumers received one €1, one €0.50, two €0.20 and one €0.10 coins such that every subsequent transaction in the MPL could be immediately implemented.

second dimension varied the information status of the shoppers. Two of the four groups, labeled n for *native*, proceeded to the MPL experiment without receiving additional information about the meaning of date marks attached to food items. The other two treatment groups, labeled e for *education*, underwent a three-step procedure to enhance their knowledge about date mark labeling: After being reminded or educated about the meanings of *BB* and *UB* labels through a tablet screen, they had to successfully complete a two-question, four-options multiple-choice quiz on date mark interpretation and had to correctly identify which of the two date mark label was present on the carton they had received. Only then were subjects in the e-treatments able to proceed to the MPL experiment.

	Information Status	
Label Type	Best-before / native	Best-before / educated
	Use-by / native	Use-by / educated

Table 2: Treatments, by treatment condition

All four treatment groups were administered the same MPL experiment, which consisted of a sequence of three screens. On each screen, the consumers made seven consecutive choices between a 'keep option' and an 'exchange option'. Choosing the former meant retaining the carton received as a reward and dated-marked with an expiry date on the day of the experiment. Choosing the latter meant exchanging it, for a financial transaction, against another, otherwise identical carton with a different expiry date. Table 3 shows the entries of the price list, with the keep option on the left and the exchange option on the right, ordered by size of payment. The upper bound of €1.00 was determined by reference to the fact that the product on offer could be purchased in the same store for a retail price between €2.19 to €2.69 with an expiry date typically 20 days ahead. The lower bound of receiving money in order to accept a carton of eggs with a longer shelf life acted as a basic check for the presence of status quo bias (Samuelson and Zeckhauser, 1988) and endowment effects (Kahneman et al., 1991). The number of days before the expiry date, x, moved from for $x = 7$ on the first to $x = 1$ (written as 'tomorrow' in the MPL) on the second screen. For $x = -1$ (written as 'yesterday') on the third screen, the exchange option lists *receive* instead of *pay* and vice versa to account for the inversion of the remaining lifetime of the product.

Keep option	Exchange option
Carton dated today	Carton dated in x days and pay €1.00
Carton dated today	Carton dated in x days and pay €0.80
Carton dated today	Carton dated in x days and pay €0.60
Carton dated today	Carton dated in x days and pay €0.40
Carton dated today	Carton dated in x days and pay €0.20
Carton dated today	Carton dated in x days and pay nothing
Carton dated today	Carton dated in x days and receive €0.20

Table 3: Price list, basic format

To maintain incentive compatibility, one of the $7 \times 3 = 21$ choices were randomly implemented for each subject. Upon conclusion of the MPL, each subject therefore either held a pack of eggs expiring on the date of the experiment and €2.00 in coins; or a pack with a different expiration date and a cash balance in accordance with the implemented choice. Then, subjects proceeded to the last stage of the experiment, the buy-back task. This task was essentially another MPL experiment in which subjects had to select, within a list of six possible alternatives, whether to keep the pack of eggs or return it to the staff in exchange for a payment (see Table 4). The buy-back stage ensures that the experimenter can observe whether participants have a negative WTP for the food item currently in their possession and interpret the evidence from the preceding MPL experiment accordingly without concerns about free disposal censoring.[20]

Keep option	Exchange option
Keep carton	Return carton and receive €3.50
Keep carton	Return carton and receive €2.50
Keep carton	Return carton and receive €1.50
Keep carton	Return carton and receive €0.50
Keep carton	Return carton and receive nothing
Keep carton	Return carton and pay €0.50

Table 4: Buy-back task

As in other MPL experiments, incentive compatibility is maintained by one of the choices being randomly selected for implementation. The experiment concluded with a brief questionnaire on demographic characteristics.

We conducted the experiment with a sample size of $n = 200$, evenly distributed across the four treatment cells based on a randomized assignment protocol. The size of the sample was determined on the basis of a power analysis informed by previous results in the literature. The closest parallel to our design is Wilson et al. (2017) who determine

[20]To detect free disposal despite the buy-back mechanism, we also monitored waste bins in the proximity of the store during the duration of the experiments. There was no evidence of cartons of eggs being disposed of close-by. Whether participants took eggs home to dispose of them is something that our design does not allow us to observe.

differences in willingness to pay for four different date mark types in a laboratory experiment. Using their effect sizes (0.41) and standard deviations (between 0.04 and 0.08) as well as conventional levels of significance ($\alpha = 0.05$) and power ($\beta = 0.8$) the most pessimistic n per treatment cell is strictly below 30 subjects. Given our field setting, we took a more conservative stance and recruited 50 subjects per cell. This allows us to detect the effect size of Wilson et al. (2017) as long as the standard deviation in an in-store setting is no more than one magnitude greater than that measured in the laboratory.

5 Hypotheses and Results

5.1 Hypotheses

The experiment is designed to detect two treatment effects. The primary effect is the impact of date mark type (*BB* vs. *UB*) on consumers' consequential decisions about perishable food.

Hypothesis 1 *Willingness to pay for perishable food is higher for products labeled best-before than for products labeled use-by.*

Theory predicts that the strongest support for hypothesis 1 will arise for expiry dates on the day of the experiment (see section 3). The reliance of hypothesis 1 on a comparison of WTP across date mark types re-emphasizes the usefulness of a buy-back experiment that generates WTP *level* data.

The same theoretical considerations predict that the date mark type will affect how consumers' valuation varies around the expiry date. The *UB* date mark, being associated with greater perishability than the *BB* mark, is predicted to lead to larger day-to-day WTP differentials around the expiry date: Consumers are expected to discount food more heavily as the items approach or exceed the end of their posted shelf-life. These *intertemporal* WTP differentials are recovered from subjects' choices in the MPL experiment and form the basis of hypothesis 2.

Hypothesis 2 *intertemporal WTP differences for perishable food are lower for products labeled best-before than for products labeled use-by.*

Both hypothesis 1 and hypothesis 2 therefore make statements about the impact of the date mark on WTP and measure these impacts around the expiry date. The difference is that hypothesis 1 draws on WTP for products expiring exactly on the day of the experiment while hypothesis 2 examines the changes around the expiry date, both before and after. Differently put, hypothesis 1 tests for differences in levels while hypothesis 2 test for differences in slopes.

The secondary effect we expect to detect through the experiment is the impact of educating (or reminding) consumers about the meaning of date marks before they take their choices in the MPL experiment. In light of prior evidence (see section 2), we expect

a significant share of consumers to be confused or mistaken about date mark meaning in the baseline condition (n). Educated consumers understand that products labeled *UB* are more perishable. Since they are expected to constitute a greater share of consumers in treatment condition (e), average WTP is predicted to be lower compared to the native treatment condition (n) for *UB* products and vice versa for *BB* products.

Following the structure of hypotheses 1 and 2, we examine this prediction both in terms of its effects on the WTP level and the intertemporal WTP differentials. For the WTP level, this leads to hypothesis 3.

Hypothesis 3 *Compared to native consumers, educated consumers have a higher WTP for perishable food labeled best-before and a lower WTP for perishable food labeled use-by.*

Again, theoretical considerations favor detection around the expiry date. Consumers' understanding of how the passing of the expiry date affects food safety and quality is likely to have the greatest impact on product valuation. The same effect is expected to establish itself in intertemporal WTP differentials.

Hypothesis 4 *Compared to native consumers, educated consumers exhibit greater intertemporal WTP differences for perishable food.*

A final consideration for guiding the data analysis is that the two-by-two design allows the experiment to also speak to the combination of the primary effect of label type and the secondary effect of education. The effects are likely to be mutually supportive, giving rise to a positive interaction effect.

5.2 Main Results

Our procedures recruited 50 subjects for each of the four treatment cell in a balanced design. All 200 subjects completed the modified MPL experiment, including the terminal buy-back mechanism. There were no drop-outs. Filters for eliminating subjects who declared themselves to be following a vegan diet or to have someone with an egg allergy in their household were present, but no subject triggered these filters in our sample.

Data from MPL experiments with general population subjects tends to be noisy and deviations from basic axioms of choice are common. Our in-store experiment is no exception. For example, between 4% and 7% of subjects switch more than once in one of the three ordered bid price lists, 15 % in at least one of the three lists. These shares are in line with or below those reported in other MPL experiments with general population subjects (Gaudecker et al., 2011). More importantly, the frequency of deviations is statistically unrelated either to treatment assignment or to expiry dates. We err on the conservative side and exclude subjects from the subsequent analysis of MPL data if they switched more than once on at least one screen, leaving us with with complete MPL observations from 170 subjects. The MPL data used for hypothesis testing includes, however, the sizeable share of non-switching subjects. While subjects that have a single

switching point always constitute the largest share, subjects exhibit considerable status quo bias. For the terminal buy-back mechanism, relevant for testing hypotheses 1 and 3, the randomized incentive mechanism returns 123 observations with an expiry date on the day of the experiment. We find that 18% of subjects have a strictly non-positive WTP for a carton, highlighting the potential scale of free disposal censoring in the absence of a terminal buy-back mechanism.

Figures 1 and 2 summarize and display the key statistics that form the basis of hypothesis testing. Figure 1 is a histogram that reports, for each treatment condition, participants' average WTP for an egg carton with an expiry date on the day of the experiment. WTP measurements for this day are most favorable for detecting the treatment effects of both date mark type and education.

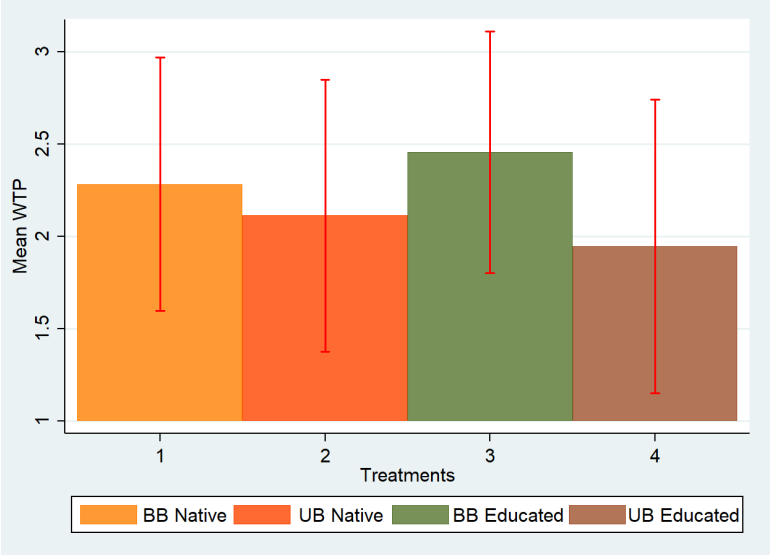

Figure 1: Average WTP in € for an egg carton with expiry date on day of experiment, by treatment condition

A first observation is that the average WTP between €1.94 and €2.46 reported by subjects are reasonable in the context of an in-store retail price for the same carton of eggs between €2.19 and €2.69. While our analysis exclusively focuses on the treatment effects, these levels indicate that subjects were, on average, taking meaningful choices in the MPL experiment. A second observation is that the standard errors, represented by the error bars around the average WTP, are relatively large: Across subjects, there is considerable variation in WTP, which reaffirms earlier findings that general population subjects tend to return noisy WTP data in MPL experiments.

Figure 2 displays, for each expiry day comparison and for all four treatment conditions, the mean intertemporal WTP differential. Comparisons are always relative to a carton of eggs expiring on the day of the experiment. The top left panel reports on the comparison with a carton of eggs with an expiry date one week after the experiment, the top right on the comparison with an expiry date one day after the experiment.

The bottom panel reports on the comparison with an expiry date one day before the experiment.

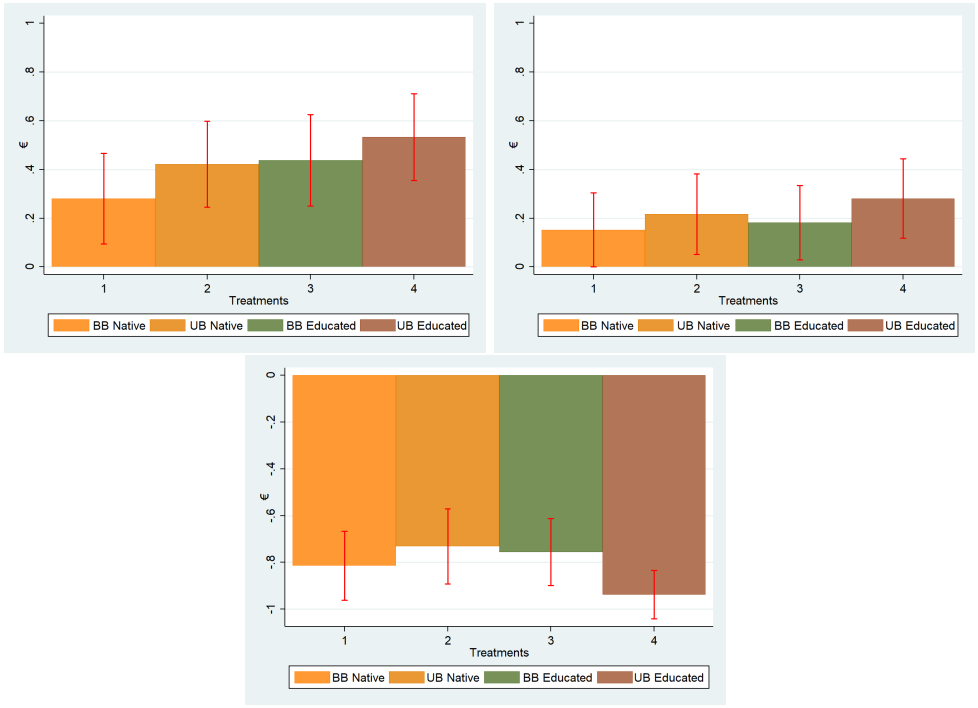

Figure 2: intertemporal WTP differential relative to carton with expiry date on day of experiment, in €, by treatment condition. Top left: Expiry date one week after; top right: one day after; bottom center: one day before experiment

Like the level estimates, the intertemporal WTP differentials pass a visual plausibility test: For eggs with remaining shelf life (top panels), the differentials are positive and somewhere between €0.15 and €0.55. The differentials are higher for one week of shelf life remaining (right) than for one day remaining (left).[21] For expired eggs, the differentials are negative, with differentials between €-0.73 and €-0.93, depending on the treatment. Like the level results, intertemporal WTP differentials exhibit a fair degree of heterogeneity, as the error bars indicate.

Result 1 reports on testing hypothesis 1 using the WTP data from the buy-back experiment as reported in figure 1.

Result 1 *The date mark type has no statistically significant effect on willingness to pay for perishable food, irrespective of information status: Willingness to pay for egg cartons date-marked* ***best-before*** *and willingness to pay for egg cartons date-marked* ***use-by*** *are statistically indistinguishable, both for native information and after consumer education.*

[21] Note the difference in scaling on the y-axes.

For both conditions of the information status, a comparison of the mean WTP indicates a negative treatment effect of selling the same perishable good under a *UB* date mark rather than a *BB* date mark. This is in line with the prediction of hypothesis 1. With native information (condition n), the difference between means of 2.28 (BB) and 2.11 (UB) is not significant, however ($p = 0.47$, M.W. Rank Sum Test). With educated consumers, the difference between means of 2.46 (BB) and 1.94 (UB) is larger, but still not significant ($p = 0.22$, M.W. Rank Sum Test). This indicates that given the heterogeneity among shoppers, the information-based policy induces some differences in consumer valuation of food, but not at a significant level. This is true even on the date of expiry, that is, in circumstances in which the information conveyed by the date mark type reaches its maximum consequentiality.

The test of hypothesis 2 is performed on the basis of the intertemporal WTP differentials recovered from subjects' choices in the MPL experiment and reported in figure 2.

Result 2 *There is a statistically significant impact of the date label on the intertemporal differences in WTP, but only for educated consumers valuing perishable products beyond the expiry date: Educated consumers have a greater WTP differential for products labeled* ***use-by*** *than for products labeled* ***best-before****.*

When consumers are educated, the difference between the mean WTP differential is only significant beyond the expiry date (-0.74 (BB) vs. -0.93 (UB), $p = 0.05$, M.W. Rank Sum Test). At all other expiry dates, the difference is not significant ($p > 0.22$ across all other dates, M.W. Rank Sum Test).

Hypothesis 3 concerns the effect of education returns on WTP. To test the hypothesis, we compare WTP level data (Figure 1) separately for each date mark type in order to detect a treatment effect of education.

Result 3 *Education has no statistically significant effect on WTP: Comparing WTP of native and educated consumers, WTP for perishable food labeled* ***best-before*** *and for perishable food labeled (****use-by****) is statistically indistinguishable.*

Under a *BB* label, mean WTP increases from €2.28 for consumers drawing on their native information to €2.46 for consumers that have undergone education about label meanings. Under a *UB* label, mean WTP decreases from €2.1 under native information to €1.94 following education. These effects are in the direction predicted by hypothesis 3, but do not reach conventional levels of statistical significance ($p = 0.49$ and $p = 0.28$ for *BB* and *UB*, respectively, M.W. Rank Sum Test). Given the heterogeneity among shoppers, education induces the desired impact on valuation of perishable food, but even on the date of expiry not at a significant level.

Result 4 *intertemporal WTP differences for perishable food differ between educated consumers and consumers drawing on native knowledge, but only for expired food products labeled* ***use-by****.*

When goods are labeled *UB* and consumers trade off an expiry date on the day of the experiment with an expiry date on the date before the experiment, the difference between the mean intertemporal WTP differential of native and educated consumers is borderline significant (-0.74 (n) vs. -0.93 (e), $p = 0.06$, M.W. Rank Sum Test). For good labeled *BB* and at all other expiry dates, the difference is not significant ($p > 0.39$ across all other dates, M.W. Rank Sum Test).

5.3 Econometric evidence

We complete the analysis of the experiment by drawing on the additional data about subjects' demographic and shopping characteristics that the survey at the outset of the experiment collected. This econometric approach additionally serves as a robustness check on our main results and allows us to formally test for interaction effects.

Tables 5 6 and 7 report on the econometric results. Table 5 explains WTP for goods with an expiry date on the day of the experiments as a function of the treatment conditions and subjects' demographic characteristics, moving from very parsimonious (only treatment conditions) to increasingly richer specifications. We use two estimation methods: A basic OLS regression on the mid-points of the buy-back intervals and a tobit estimation. The baseline across all methods and specifications is the treatment condition *BB/native*.

	OLS			Tobit		
	(1)	(2)	(3)	(4)	(5)	(6)
UB label	-.330	-.170	.0221	-.418	-.333	.0376
	(.347)	(.489)	(.520)	(.600)	(.855)	(.169)
Education	.008	.172	.323	.0515	.138	.436**
	(.346)	(.460)	(.520)	(.600)	(.828)	(.168)
UB x Education		-.339	-.644		-.178	-.893***
		(.696)	(.791)		(1.201)	(.228)
Constant	2.357***	2.284***	2.179***	2.980***	2.941***	2.74***
	(.293)	(.335)	(.632)	(.535)	(.619)	(.156)
Socio-dem. controls	No	No	Yes	No	No	Yes

Table 5: Dependent variable: WTP at expiry date. Baseline: Date mark *BB* at *n*ative information, no socio-demographic controls. Standard errors clustered at individual level. $^*p > 0.1$, $^{**}p > 0.05$, $^{***}p > 0.01$

Irrespective of method and specifications, the coefficient of the *UB* date mark type is negative, as predicted, but never attains statistical significance. Once socio-demographic controls are included (specifications (3) and (6)), the quantitative relevance of the *UB* mark decreases by an order of magnitude. The coefficient of the *e*ducation treatment has the predicted positive effect on WTP and becomes quantitatively more relevant by including socio-demographic controls, leading to a statistically significant effect on WTP in specification (6): When sociodemographic controls are included, a tobit estimation, capable of accounting for the censored nature of our data, predicts an increase of €0.44 in WTP for goods labeled *BB* following education. In the same specification, we also recover a negative and highly significant interaction effect between a *UB* label and

education. The magnitude of the interaction effect more than offsets the education effect on WTP under a *BB* label alone. This provides further evidence that jointly, education and the *UB* mark decrease WTP, thus adding robustness to the non-parametric tests of hypotheses 1 and 3.

Table 6 explains the intertemporal WTP differentials measured in the experiment as a function of remaining shelf-life, treatment condition, and subjects' characteristics, again moving from very parsimonious to increasingly richer specifications. We again use two estimation methods: A basic OLS regression on the mid-points of the MPL intervals, and a tobit estimation. The baseline across all methods and specifications is the treatment condition *BB/native* with a remaining shelf-life of one week and no socio-demographic controls.

	OLS			Tobit		
	(1)	(2)	(3)	(4)	(5)	(6)
Expires tomorrow	-.208***	-.208***	-.208***	-.295***	-.295***	-.292***
	(.035)	(.035)	(.036)	(.053)	(.053)	(.053)
Expired yesterday	-1.228***	-1.228***	-1.228***	-1.840***	-1.841***	-1.826***
	(.059)	(.059)	(.060)	(.134)	(.134)	(.131)
UB label	.038	.081	.073	.043	.134	.123
	(.058)	(.080)	(.077)	(.091)	(.125)	(.117)
Education	.044	.085	.089	.063	.150	.151
	(.058)	(.080)	(.077)	(.091)	(.125)	(.116)
UB x Education		-.084	-.066		-.180	-.140
		(.116)	(.108)		(.182)	(.167)
Constant	.374***	.354***	.398***	.531***	.488***	.566***
	(.059)	(.064)	(.102)	(.095)	(.101)	(.164)
Socio-dem. controls	No	No	Yes	No	No	Yes

Table 6: Dependent variable: WTP differential between expiry date on day of experiment and expiry date in one week. Baseline: Date mark *BB* with *n*ative information, no socio-demographic controls. Standard errors clustered at individual level. $^{*}p > 0.1$, $^{**}p > 0.05$, $^{***}p > 0.01$

Across specifications, the coefficient estimates for the pooled data show that remaining shelf life consistently matters for WTP. Reducing remaining shelf life from one week to one day decreases WTP by €0.21 to €0.30. Expired food items are valued between €1.23 and €1.84 less. These effects are all statistically significant at the 1% level. But neither a *UB* date mark nor education have a statistically significant impact on the WTP differential. Likewise, the interaction effect between the *UB* and *e*ducation treatments is still negative, as observed in the non-parametric tests, but no longer attains statistical significance, even when socio-demographic controls are included. This provides further evidence for the findings from testing hypotheses 2 and 4 that label type and education have little detectable impact on the valuation of food items.

	OLS				Tobit	
	(1)	(2)	(3)	(4)	(5)	(6)
UB label	-.054	.083	.096	-.143	.214	.268
	(.070)	(.108)	(.102)	(.183)	(.262)	(.235)
Education	-.070	.058	.068	-.129	.205	.237
	(.070)	(.102)	(.100)	(.183)	(.250)	(.219)
UB x Education		-.265*	-.233		-.712**	-.581*
		(.139)	(.142)		(.360)	(.345)
Constant	-.749***	-.814***	-.953***	-1.260***	-1.423***	-1.530***
	(.063)	(.073)	(.111)	(.175)	(.204)	(.301)
Socio-dem. controls	No	No	Yes	No	No	Yes

Table 7: Dependent variable: WTP differential between eggs expiring on day of experiment and expired eggs. Baseline: Date mark *BB* with *n*ative information, no socio-demographic controls. Standard errors clustered at individual level. $^{*}p > 0.1$, $^{**}p > 0.05$, $^{***}p > 0.01$

Restricting the pooled sample to observations where the intertemporal comparison only concerns food items just before and after expiry, table 7 reports again no significant treatment effects of the *UB* label or education, but a significant and negative treatment effect for their interaction: For consumers educated about the meaning of date marks, WTP for food labeled *UB* is significantly lower. Adding socio-demographic controls attenuates the statistical significance of this result somewhat, but the results of the parametric analysis remain consistent with the non-parametric results. Taken together, they suggest that educating consumers about the meaning of date marks makes their valuation of food more responsive to the safety message of *use-by* labels. This provides a pathway through which education could lead to more unsafe food being discarded. At the same time, education does not make consumers' valuation more responsive to the quality message of *best-before* labels and therefore does not provide a pathway toward less food being discarded that is still of good quality.

6 Discussion and further experimental evidence

Results 1 through 4 and the econometric evidence raise doubts about the ability of information-based policies to impact consumers' behavior as intended. While the general effects on the formation and evolution of WTP go in the desired direction, the treatment effects affect an insufficient share of subjects in their respective groups consistently enough to result in impacts of statistical significance. Most problematic for food waste policies is the significant impact on WTP levels of the interaction effect of labeling food with a *UB* date mark and educating consumers: This provides some evidence that educating consumers possibly results in an increase in health-related discarding of food, but not decrease food waste.

To look for a possible explanation behind this evidence, we designed and conducted a follow-up experiment. The experiment builds on the hypothesis that information-based policies using date marks fail because shoppers pay little attention to the information conveyed through this vehicle. By design, it allows us to test of whether shoppers notice differences in date mark type (*BB* versus *UB*) and expiry dates in a choice-relevant

setting. We summarize its design and results before using its evidence to re-evaluate our main results.

The experiment was conducted in the same in-store setting as the MPL experiment.[22] Shoppers passing through the store lobby were approached and asked whether they would participate in a tablet-based survey, with a carton of free-range eggs as a reward. Filter questions eliminated shoppers that do not purchase eggs for dietary reasons. Shoppers completed a survey on shopping habits and food preferences before being offered their reward. Experimental treatments began at this step in the procedures and differed with respect to the rewards on offer. Every participant was offered two cartons of free-range eggs, one of which they could pick as their reward. In one treatment condition, labeled *mark*, the two cartons differed by date mark (*BB* versus *UB*) and by brand (*A* vs. *B*)[23], randomly mixed. In the other treatment condition, labeled *date*, the two cartons differed by expiry date (*day of the experiment* vs. *day after the experiment*) and by brand (as above), randomly mixed. The associations between the two brands and the treatment dimensions were additionally randomized, resulting in two *combinations*, 1 and 2. Before selecting their preferred carton, the survey instrument asked shoppers about the criteria that they use when buying eggs in a supermarket and asked them to apply the same criteria to their choice in the experiment. After choosing, participants proceeded to a structured interview about the reasons for their choice. The structure was sequential: The interview progressed from open-ended to increasingly targeted questions about possible differences in the date mark attached to the two cartons and stopped as soon as a subject correctly pointed out the difference. This allowed subjects to be classified in one of four groups: Those who were attentive to the date mark and for whom it was choice relevant; those who were attentive, but for whom it was choice-irrelevant; those who could be guided to be attentive and were therefore theoretically reachable by a date mark; and finally those who remained inattentive to date marks even after guidance by the experimenter.[24] The interview concluded with participants being given the option of revising their choice.

A sample of 160 individuals was assigned to the two treatments (*mark* versus *date*) and, within each treatment, to one of the two brand-treatment combinations (1 or 2). The assignment was based on a randomized assignment protocol that produced a balanced assignment of 80 in each treatment and 40 for each brand-treatment combination. Table 8 displays the results of the experiment.

[22] See the appendix for the flowchart of this experiment.

[23] We selected two brands of free-range eggs, REWE and Heitlinger, that retailed in the same price band in order to minimize strong brand-specific differences.

[24] Specifically, subjects who mentioned the difference between the date marks on the cartons without prompting by the experimenter were classified as 'choice relevant' if they mentioned the difference as the reason for their choice and 'choice irrelevant' if not. Subjects who recognized the difference after being guided to the date mark were classified as 'reachable', that is attentive after additional effort by the experimenter. All other subjects who concluded the guidance without noting the difference between the date marks were classified as 'inattentive' to date mark labels.

	Mark		Date		
	Comb. 1	Comb. 2	Comb. 1	Comb. 2	Total
Number of subjects	40	40	40	40	160
Choice relevant	0	0	6	3	9
Choice irrelevant	0	0	6	2	8
Reachable	0	1	10	11	22
Inattentive	40	39	18	24	121
UB≠BB (share)	0.85	0.73	0.73	0.8	0.78
UB correct meaning (share)	0.55	0.43	0.6	0.38	0.49
BB correct meaning (share)	0.95	0.83	0.9	0.8	0.87
Choice revised	1	2	2	1	6

Table 8: Follow-up experiment - summary statistics

The experimental evidence points to a number of observations that help interpret the results of the MPL experiment. First, as expected, brand-treatment combinations do not matter (combination 1 vs. combination 2 for mark and date treatments, $p > 0.81$, chi-square test). Second, pooling the data across brand-treatment combinations, shoppers are much more inattentive about the date mark type than about the expiry date (79 versus 42, $p < 0.01$, chi-square test). Third, irrespective of treatment, there is a significant difference between the *use-by* and *best-before* label: While 78% of subjects agree that the two labels have different and distinct meanings, only 49% correctly interpret the former as opposed to 89% for the latter ($p < 0.01$, chi-squared test).[25]

The evidence from the follow-up experiment suggests that some explanations for the results of the modified MPL experiment are more plausible than others. The ineffectiveness of date mark type manipulations on their own (result 1) is likely a result of consumers' inattentiveness to what is a small visual change in the date mark. Choosing more salient visual cues could be a remedy for this inattentiveness. The information about date marks, which is also visually more prominent, has a larger impact and is therefore choice relevant, as the intertemporal WTP differentials illustrate (result 2). The most important insight from the follow-up experiment is about result 3: Contrary to the assumptions behind some food waste policies (EU-Council, 2016), consumers are already relatively well informed about the meaning of *best-before* date marks. As a result, it is unsurprising that the education treatment causes no measurable change in willingness to pay in the BB condition. Education, however, does impact on WTP in the UB condition, however, since a substantial share of shoppers misinterpret the *use-by* date mark.

Our combined evidence has problematic implications for information-based policies targeting food waste through educating consumers as envisaged by policy-makers (EU-Council, 2016).[26] By educating the substantial share of consumers that are uninformed

[25] More evidence for misinterpretation of the *UB* date mark comes from those individuals who accepted to revise their choice: Without exception, these consumers opted for a carton labeled *UB* in exchange for a carton labeled *BB*.

[26] A current example is the UK Food Standard video campaign about the differences in date mark types: `https://www.youtube.com/watch?v=CDIpDupYPiY`.

about the meaning of *use-by* labels, such campaigns are likely to further health objectives. These objectives will be served by more food being discarded for food safety reasons. Our results predict little to no counteracting effect on food conservation in households. This suggests that food waste requires a different route of intervention.

7 Concluding Remarks

Regulations that aim to inform consumers on important features and consequences of their purchase and management decisions about perishable goods are timely. Given the wider impacts of producing, harvesting, and delivering food to consumers' homes, helping consumers to avoid food waste while ensuring public health deserves policy-makers' attention. Date-marks are a plausible approach to providing this help, and initiatives to educate consumers about how to act based on date-marks merit careful assessment.

One challenge for coming to a better assessment of date-marking policies is the nature of the evidence base from which researchers and policy-makers can argue. Previous research has made substantial contributions towards building this evidence base through large-scale surveys, vignette experiments, and laboratory studies. Due to a number of empirical challenges, causal evidence has been more difficult to obtain in order to speak to the question of how date marks types and education causally affect consumers' valuation of perishable food. Our present paper is an attempt to provide such causal evidence for the academic and policy discussion.

In our mind, the evidence presented in this paper makes two contributions. One is the key finding that date mark types, as used at present, are not impactful for consumer behavior and that educating consumers about their meaning is conducive to the health and safety objectives implicit in date marking food, but not to the objective of preventing food waste. Our follow-up experiment suggests that the explanation for this phenomenon lies in the existing asymmetry in consumers' understanding of the two date marks currently in use.

The other contribution is the nature of the evidence that our paper presents. The evidence emerges from the combination of three interlinked building blocks. The first is a preliminary survey, which informs the choice of a perishable food item that lends itself to the experimental manipulation of date marks without compromising research ethics or the safety of participants. With the help of this choice, we believe that we overcome a core challenge for causal identification, namely varying the date mark without varying the underlying product. The second building block is a modified multiple price list experiment that exchanges the laboratory setting and its student subjects for an in-store setting with members of the general population. These procedural choices require an experimental interface that is accessible for a wide range of shoppers coming to the store. It also requires a new design approach in the form of a buy-back mechanism since we cannot rely on protocol compliance to overcome free disposal censoring. The third building block is a follow-up experiment that pursues the hypothesis that the treatment effects plausibly result from the nature of demand-side information. Jointly,

these building blocks demonstrate that it is feasible to overcome at least some of the empirical challenges of generating causal evidence to understand information-based policies. Moreover, they demonstrate that the policy proposals are unlikely to accomplish their desired objectives. And they provide at least one explanation why these proposals are likely to underperform, thus informing the development of policy alternatives.

The present paper provides conceptual and methodological points of departure for future research. Much of this paper hinges on the question of how well its evidence generalizes to other food categories. Future work will need to explore this as well as the question of how this approach could be extended to other categories of perishable products. This exploration should also consider the non-food domain, where issues of waste and (planned) obsolescence raise similar issues. While we believe that the in-store experiment takes important steps towards enhanced external validity of the findings, our attempts to conduct this research as a fully-fledged natural field experiment were unsuccessful. We remain hopeful that this avenue can be embarked upon in the future, while respecting the requirements of safety and ethics. Finally, the paper is limited to testing how current date marking practices perform. Clearly, we cannot rule out, and would in fact strongly suspect, that there exist alternative information-based policies that can make a significant contribution towards increasing consumers' valuation of perishable, but safe foods and reduce food waste. Identifying these alternatives remains a research priority.

References

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2007). Valuation using multiple price list formats. *Applied Economics*, 39(6):675–682.

Collart, A. and Interis, M. (2018). Consumer imperfect information in the market for expired and nearly expired foods and implications for reducing food waste. *Sustainability*, 10(11):3835.

Coursey, D. L., Hovis, J. L., and Schulze, W. D. (1987). The disparity between willingness to accept and willingness to pay measures of value. *The Quarterly Journal of Economics*, 102(3):679–690.

Daniel, C. (2016). Economic constraints on taste formation and the true cost of healthy eating. *Social Science & Medicine*, 148:34–41.

Davis, D. and Millner, E. (2005). Rebates, matches, and consumer behavior. *Southern Economic Journal*, 72:410–421.

Ek, C. and Miliute-Plepiene, J. (2018). Behavioral spillovers from food-waste collection in swedish municipalities. *Journal of Environmental Economics and Management*, 89:168–186.

Ellison, B. and Lusk, J. L. (2018). Examining household food waste decisions: A vignette approach. *Applied Economic Perspectives and Policy*, 40(4):613–631.

EU-Council (2016). Council conclusions on food losses and food waste, june 28, 2016. Technical report, EU Council.

Eurobarometer (2015). Flash Eurobarometer 425: Food waste and date marking. Technical report, European Commission.

Farr-Wharton, G., Foth, M., and Choi, J. H.-J. (2014). Identifying factors that promote consumer behaviours causing expired domestic food waste. *Journal of Consumer Behaviour*, 13(6):393–402.

Gaudecker, H.-M. v., Soest, A. v., and Wengstrom, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, 101(2):664–94.

Hall, K. D., Guo, J., Dore, M., and Chow, C. C. (2009). The progressive increase of food waste in america and its environmental impact. *PloS one*, 4(11):e7940.

Hayes, D. J., Shogren, J. F., Shin, S. Y., and Kliebenstein, J. B. (1995). Valuing food safety in experimental auction markets. *American Journal of Agricultural Economics*, 77(1):40–53.

Hebrok, M. and Boks, C. (2017). Household food waste: Drivers and potential intervention points for design–an extensive review. *Journal of Cleaner Production*, 151:380–392.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1):193–206.

Lebreton, M., Langdon, S., Slieker, M., Nooitgedacht, J., Goudriaan, A., Denys, D., van Holst, R., and Luigjes, J. (2018). Two sides of the same coin: Monetary incentives concurrently improve and bias confidence judgment. *Science Advances*, 4:eaaq0668.

Murad, Z., Sefton, M., and Starmer, C. (2016). How do risk attitudes affect measured confidence? *Journal of Risk and Uncertainty*, 52(1):21–46.

Neff, R. A., Spiker, M., Rice, C., Schklair, A., Greenberg, S., and Leib, E. B. (2019). Misunderstood food date labels and reported food discards: A survey of u.s. consumer attitudes and

behaviors. *Waste management*, 86:123–132.

Neff, R. A., Spiker, M. L., and Truant, P. L. (2015). Wasted food: U.s. consumers' reported awareness, attitudes, and behaviors. *PloS one*, 10(6):e0127881.

Quested, T. E., Marsh, E., Stunell, D., and Parry, A. D. (2013). Spaghetti soup: The complex world of food waste behaviours. *Resources, Conservation and Recycling*, 79:43–51.

Reinhardt, P. G. et al. (1973). A theory of household grocery inventory holdings. *Kyklos*, 26(3):497–511.

Roe, B. E., Phinney, D. M., Simons, C. T., Badiger, A. S., Bender, K. E., and Heldman, D. R. (2018). Discard intentions are lower for milk presented in containers without date labels. *Food quality and preference*, 66:13–18.

Samuelson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1):7–59.

Secondi, L., Principato, L., and Laureti, T. (2015). Household food waste behaviour in eu-27 countries: A multilevel analysis. *Food policy*, 56:25–40.

Shogren, J. F., Shin, S. Y., Hayes, D. J., and Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *The American Economic Review*, pages 255–270.

Siddique, M. A. M. (2012). Explaining the role of perceived risk, knowledge, price, and cost in dry fish consumption within the theory of planned behavior. *Journal of global marketing*, 25(4):181–201.

Stenmarck, Â., Jensen, C., Quested, T., Moates, G., Buksti, M., Cseh, B., Juul, S., Parry, A., Politano, A., Redlingshofer, B., et al. (2016). *Estimates of European food waste levels*. IVL Swedish Environmental Research Institute.

Thompson, B., Toma, L., Barnes, A. P., and Revoredo-Giha, C. (2018). The effect of date labels on willingness to consume dairy products: Implications for food waste reduction. *Waste management*, 78:124–134.

Tsiros, M. and Heilman, C. M. (2005). The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. *Journal of marketing*, 69(2):114–129.

Wilson, N. L., Miao, R., and Weis, C. (2018). Seeing is not believing: Perceptions of date labels over food and attributes. *Journal of food products marketing*, 24(5):611–631.

Wilson, N. L., Rickard, B. J., Saputo, R., and Ho, S.-T. (2017). Food waste: The role of date labels, package size, and product category. *Food Quality and Preference*, 55:35–44.

Yu, Y. and Jaenicke, E. C. (2020). Estimating food waste as household production inefficiency. *American Journal of Agricultural Economics*, 102(2):525–547.

A Appendix – Survey questions

General Survey (pt. 1)

1. Do you regularly buy groceries?
    - Yes
    - No
    - I prefer not to answer
2. Would you saying that you are doing the most grocery shopping of your household?
    - Yes
    - No
    - I prefer not to answer
3. How many people, including yourself, live in your household?
    - 1
    - 2
    - 3
    - 4
    - 5 or more
    - I prefer not to answer
4. *If more than one person in the household:* Is someone in your household vegetarian?
    - Yes
    - No
    - I prefer not to answer
5. Are you vegetarian?
    - Yes
    - No
    - I prefer not to answer
6. *If more than one person in the household:* Is someone in your household vegan?
    - Yes
    - No
    - I prefer not to answer
7. Are you vegan?
    - Yes
    - No
    - I prefer not to answer
8. If we may ask: Could you assign yourself to one of the following age groups?
    - 18-25
    - 26-35
    - 36-45
    - 46-55
    - 56-65

- 66 or more
- I prefer not to answer

9. Do you have children?
   - Yes
   - No
   - I prefer not to answer
10. Which gender do you associate with?
    - Female
    - Male
    - Other
    - I prefer not to answer
11. Which gender do you associate with?
    - Female
    - Male
    - Other
    - I prefer not to answer
12. What is the highest educational attainment you have achieved?
    - *Kein Schulabschluss* - no High School Diploma.
    - *Hauptschulabschluss* - Hauptschule Diploma, awarded after 9 years of Education.
    - *Realschulabschluss* - Realschule Diploma, awarded after 10 years of Education.
    - *Abitur* - High School Diploma, awarded after 12 years of Education and necessary prerequisite for University.
    - *Berufsschulabschluss* - Apprenticeship, awarded after 3-years education cycle in which the student can learn a craft work from a trainer (an expert in the field) and attend a vocational school.
    - *Hochschulabschluss* - University Degree, with no difference whether it is Bachelor or Master.
    - *Promotion* - Ph.D.
13. You are currently...
    - Studying/Training
    - Self-employed
    - Employed
    - Retired
    - Unemployed
    - I prefer not to answer.

General Survey (pt. 2)

N.B. The second part of the general survey is only administered to subjects in the modified MPL experiment.

1. What do you usually pay most attention to when you buy groceries at the grocery store:
    - Produced in Germany
    - Sustainable or environmentally friendly production
    - Price
    - Brand
    - Other [please specify].
2. Do you or anyone else in your household like to eat eggs?
    - Yes
    - No
3. Do you or anyone else in your household have an egg allergy?
    - Yes
    - No
4. Is food sometimes thrown away in your household?
    - Yes
    - No
5. Who is typically responsible for discarding such food?
    - Me
    - Someone else

B Flowcharts of experimental designs

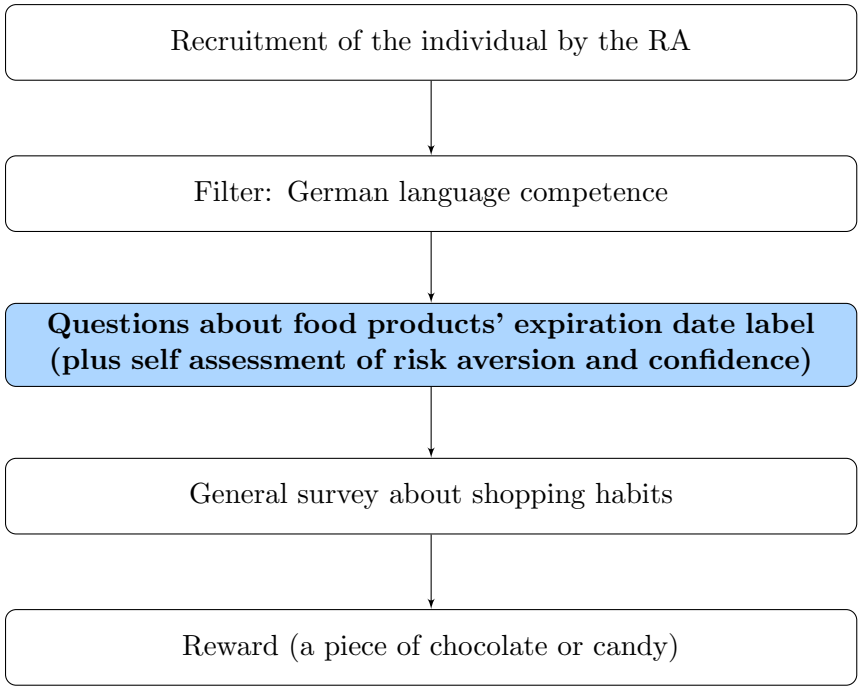

Figure A1: Flowchart of the preparatory survey

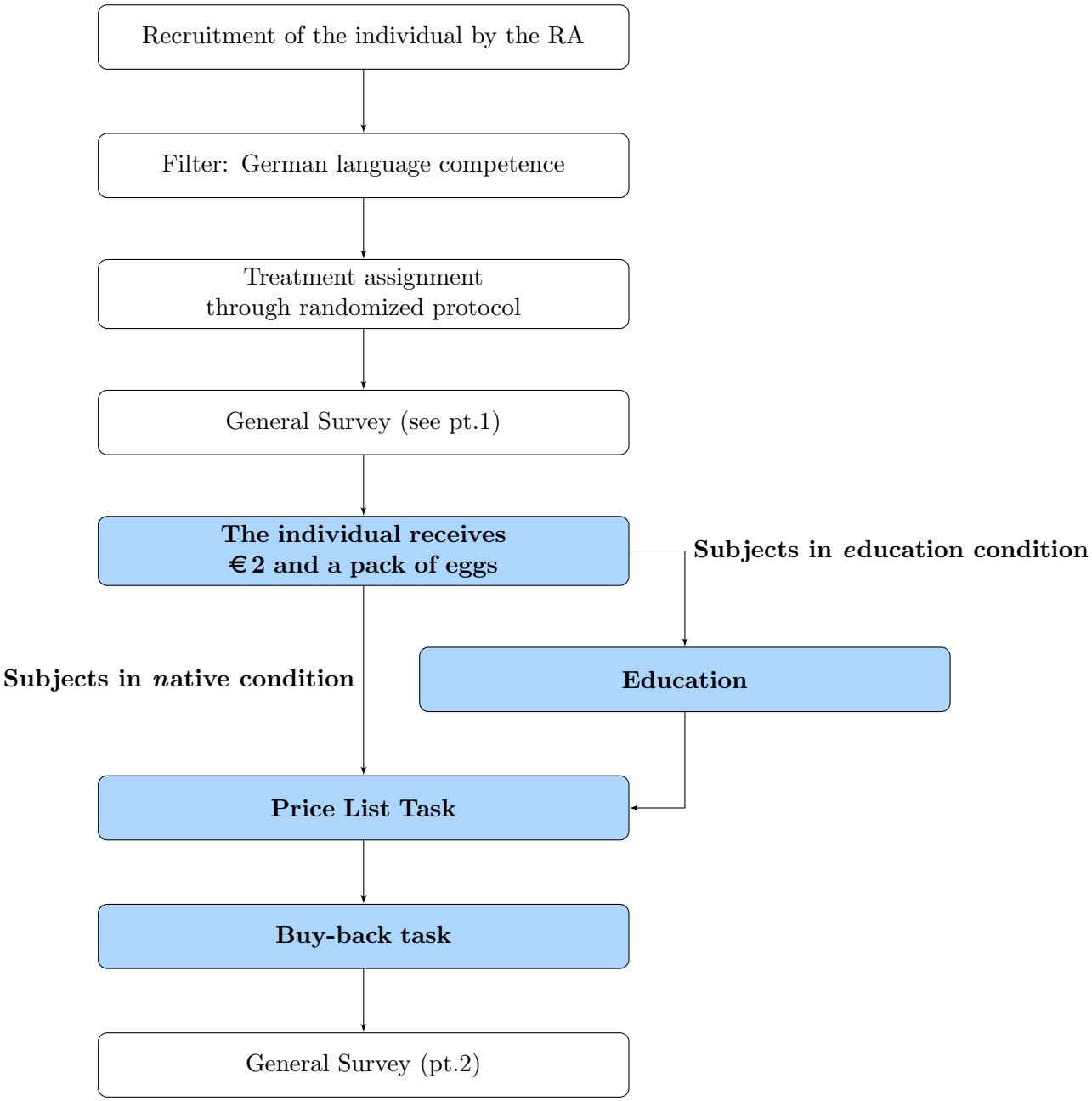

Figure A2: Flowchart of the modified MPL Experiment

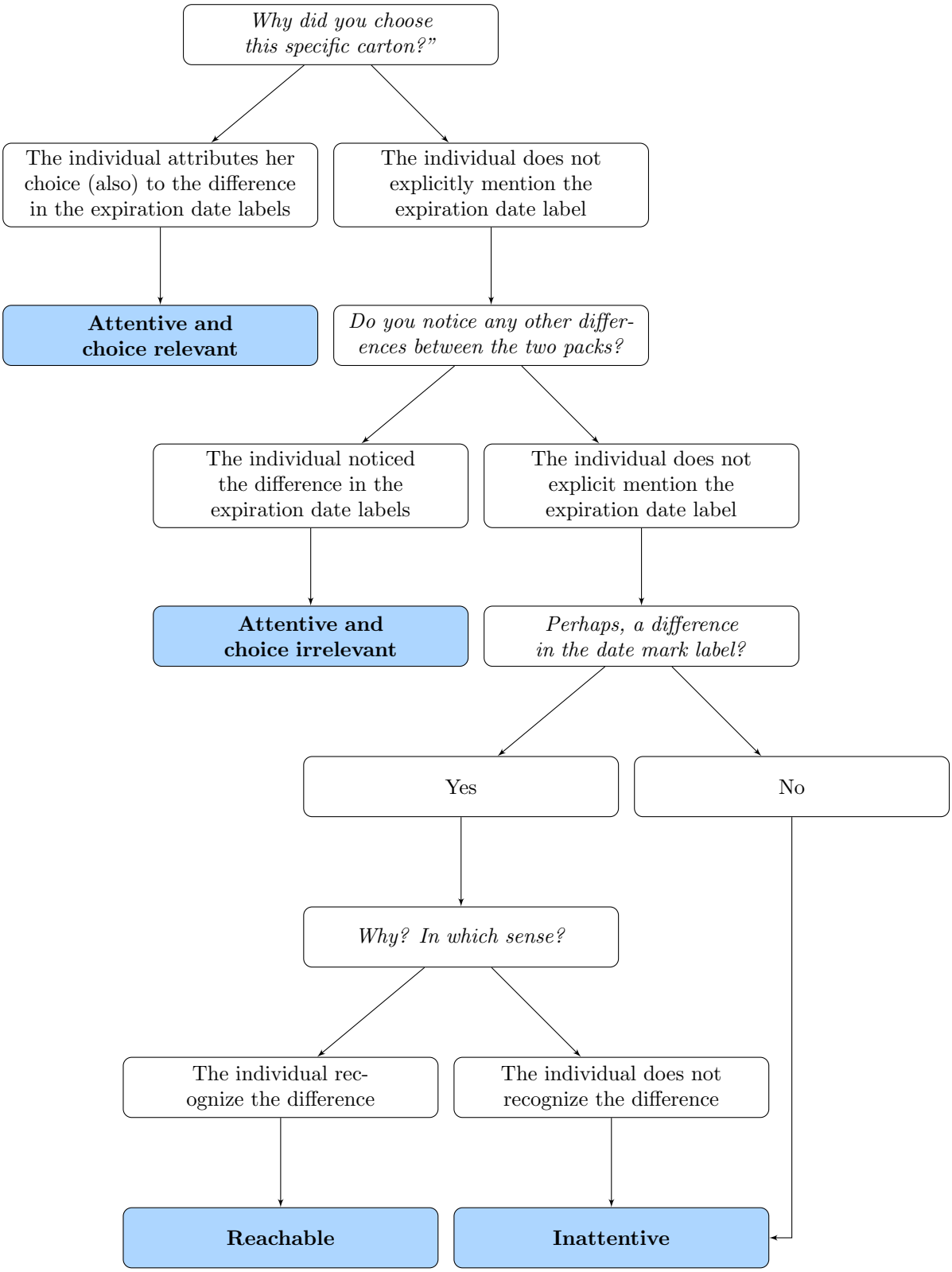

Figure A4: Flowchart of the Attention Test (Inattention Experiment)

The questions for the Information test are as follows

1. *In shops, packaged groceries are sold with two labels indicating expiration dates. One is "Best Before", the other is "Use By". What do you think:*
    - The two date marks have different messages for the consumer.
    - The two date marks have the same meaning for the consumer.
2. *The "use by" date means:*
    - The food will be safe to eat up to this date and should not be eaten past this date (UB)
    - The food can be consumed after this date, but it may no longer be at its best quality (BB)
    - The food can be used after this date only if the packaging is not damaged (CD1)
    - The food will be safe to eat from this date on and should be eaten past this date (CD2)
3. *The "best before" date means:*
    - The food will be safe to eat up to this date and should not be eaten past this date (UB)
    - The food can be consumed after this date, but it may no longer be at its best quality (BB)
    - The food can be used after this date only if the packaging is not damaged (CD1)
    - The food will be safe to eat from this date on and should be eaten past this date (CD2)

The order of the two questions and the order of the possible options (equal in the two question for each individual in order to avoid possible confusion) is random across the subjects. The answers used in these two questions are the definition of the UB label and of the BB and two confounding definitions (CDs). The label definitions in English and in German are the official ones used in the Eurobarometer (2015).

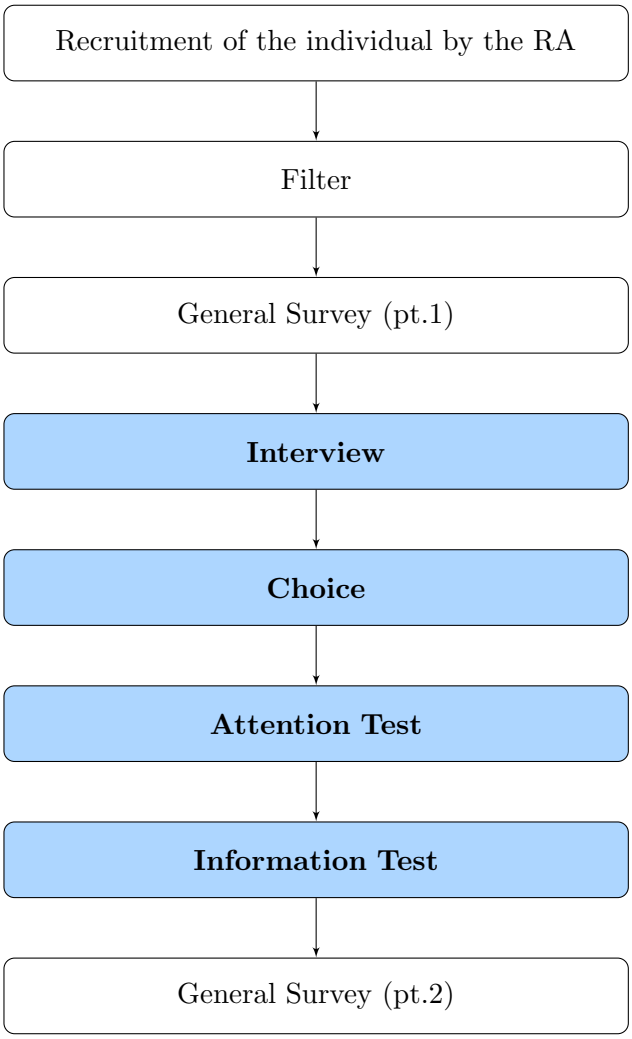

Figure A3: Flowchart of the Attentiveness Experiment