



Behavioural economics

D4.4 Linking Bayesian and agent-based models to assess consumer food waste



Authors

Matthew Grainger, University of Newcastle, UK

Simone Piras, University of Bologna, Italy

Simone Righi, University of Bologna, Italy

Marco Setti, University of Bologna, Italy

Gavin Stewart, University of Newcastle, UK

Matteo Vittuari, University of Bologna, Italy

Project coordination and editing provided by Ecologic Institute.

Manuscript completed in April 2019

Document title Integrated model of consumer behaviours in relation to food waste

Work Package WP4

Document Type Deliverable

Date 30/04/2019

ISBN 978-94-6343-994-7

DOI <https://doi.org/10.18174/477976>

Acknowledgments & Disclaimer

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641933.

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the publisher is given prior notice and sent a copy.

Table of Contents

1	Executive summary	6
2	Introduction and objectives	7
3	Theoretical background	8
4	The integrated model	12
	4.1 The Agent-Based Model	13
	4.2 The Bayesian Network	14
	4.3 The integrated model	15
5	Calibration and basic interventions	16
	5.1 Creation of simulated populations	16
	5.2 Changes to baseline to assess impact on food waste	17
6	Results and discussion	20
	6.1 Results of the simulations: baseline and changes	20
	6.2 Discussion of the results of preliminary simulations	30
7	Conclusions and further applications	34
8	References	36
9	Annex	37
	9.1 Technical details of the individual model components	37
	9.1.1 Consumer Agent Based Models	37
	9.1.2 Consumer Bayesian Network Model	40
	9.2 Data structure	40
	9.3 MatLab and R codes	43

List of Tables

<i>Table 1: Elements of the integrated ABM-BN model (variables).</i>	9
<i>Table 2: Food waste classes for the extraction of individual values.</i>	14
<i>Table 3: Composition of the sample in terms of family size and ages.</i>	16
<i>Table 4: Sample of matrix uploaded on MatLab ("awareness" of food waste effects).</i>	17
<i>Table 5: Regression coefficients between each motivation and all the others.</i>	17
<i>Table 6: Relative "importance" of the variables in 1000 neural networks.</i>	32
<i>Table 7: Synthetic statistical indicators for all calibration variables.</i>	41
<i>Table 8: Distribution of the answers for all calibration variables (%).</i>	42

List of Figures

<i>Figure 1: WP1 REFRESH Food Waste Model</i>	9
<i>Figure 2: Scheme of the interrelationships within the integrated ABM-BN model.</i>	12
<i>Figure 3: Graphical representation of the BN used in the integrated ABM-BN model.</i>	15
<i>Figure 4: Self-declared frequency of actions (1-2). Errore. Il segnalibro non è definito.</i>	
<i>Figure 5: Self-declared frequency of actions (3-10).</i>	22
<i>Figure 6: Self-declared frequency of actions (11-18).</i>	23
<i>Figure 7: Self-declared frequency of actions (19-26).</i>	24
<i>Figure 8: Food waste across time (mean and median) in the baseline setting.</i>	25
<i>Figure 9: Food waste distribution at times 1 (WP1 data) and t_{max}(baseline).</i>	26
<i>Figure 10: Standard deviation of mean waste across populations (baseline).</i>	26
<i>Figure 11: Evolution of "awareness" and "attitude" along time (baseline).</i>	27
<i>Figure 12: Evolution of competing goals related to food across time (baseline).</i>	28
<i>Figure 13: Reduced machine-learnt BN of Dutch data from REFRESH WP1 survey.</i>	33

List of abbreviations

ABM	Agent-Based Model
BN	Bayesian Network
EU	European Union
FWM	Food Waste Model
FWP	Food Waste Prevention
WP	Work Package

Glossary

Agents	Actor and decision makers in a model.
Agent Based Model	A class of computational models for simulating the actions and interactions of autonomous agents (either individual or collective entities such as organizations or groups) with a view to assessing their effects on the system as a whole.
Bayesian Networks	Probabilistic graphical models (a type of statistical model) that represent a set of random variables and their inter-dependencies (i.e. conditional dependencies). These graphical structures are used to represent knowledge about an uncertain domain. Bayesian Networks consist of "nodes" (variables) and "arcs" (conditional links represented by arrows).
Behavioural economics	Study of the effects of psychological, social, cognitive, and emotional factors on the economic decisions of individuals and institutions and of their consequences for market prices, returns, and resource allocation.
Impact analysis	Examination and measure of the effects of an event on a system or part of it in a specified area.
Machine learnt/ learning	Machine learning is the use of mathematical algorithms (sets of rules) to help identify patterns in data. A machine learnt structure of a Bayesian Network is where the computer chooses the most parsimonious structure (the least complex structure to explain the variation in the data) regardless of theoretical constraints.
Model structure	Statistical models are made up of one or more dependent variables that are related in some way to one or more independent variables. The dependent variable(s) is/are the one(s) which we wish to understand or predict from the independent variables. Model structure (how all the independent variables are arranged and how they

are related to each other) is an important but often overlooked aspect of modelling.

Roadmap A plan that matches short-term and long-term goals with specific solutions to help meeting those goals. In the case of REFRESH, the roadmap aims at identifying the major steps to be undertaken to reach a reduction of 50% of European food waste by 2030. The roadmap is based on the results of the simulation of the integrated Agent Based Model and Bayesian Network.

Scenario analysis A process of analysing possible future events by considering alternative possible outcomes. In the case of REFRESH WP4, it is based on the results of the simulation of the integrated Agent Based Model and Bayesian Network.

Socio Economic model Examines and depicts social, ethic and economic factors to better understand how their combination influences consumer behaviour and shapes the economy.

1 Executive summary

This report presents the results of the first testing and simulations from an integrated model of consumer food waste. This research is the first step in developing a model that can assess the impact of policy interventions on reducing food waste at the consumer level. The goal is to create the foundation for a predictive, decision-relevant, and dynamic policy support tool, which can provide a road map for the 50% reduction of European food waste by 2030. This integrated model combines two approaches, an Agent Based Model and a Bayesian Network. Integrating these two model types allows researchers to include interpersonal dynamics when assessing the impact of policy interventions such as information campaigns, economic incentives or trainings. As the current work represents an overview of the first testing and simulations, this report is primarily aimed at researchers developing behavioural economic models also to stimulate further advancements in this domain. At the same time, the more mature versions of this model are intended to represent a support tool for decision-makers by providing them with estimates about potential hotspots for change (e.g. estimates of the amount of food waste potentially generated by consumers belonging to different age groups, thus suggesting the groups that should represent a priority for awareness campaigns) and changes that might be introduced by certain interventions.

2 Introduction and objectives

REFRESH is an EU research project dedicated to contributing to the achievement of Target 3 of Sustainable Development Goal 12, i.e. to halve per capita food waste at the retail and consumer level as well as reducing food losses along the food chain by 2030. Partners across Europe are collecting data on methods to reduce or re-purpose food waste.

In developed countries an estimated 30 to 40% of food is wasted; about half of this waste stems from consumers, while the remainder is lost through farm practices, transport and processing, as well as in a retail setting (Godfray et al. 2010; Gustavsson et al. 2011). To meet target 12.3 of the Sustainable Development Goals of “halving per capita food waste and reducing food losses by 2030”, there needs to be a better understanding of the drivers of consumers’ food waste (as the largest single contributor to food waste). More importantly, the effectiveness of interventions designed to reduce food waste at consumer level needs to be assessed.

Using a simulation approach is necessary for assessing food waste, because empirical data are often very limited and low quality (Xue et al. 2017). Where data are available, they have a high potential for bias (like in the case of self-reported consumer food waste) or are limited in scale. This leads to high levels of uncertainty in the available data, additional to the complexity associated with understanding the socio-economic drivers of food waste. Bayesian Networks (BNs) can incorporate uncertainty and complexity in the model structure but are less effective at incorporating behavioural factors (i.e. specific biases of single actors, and interactions among actors) and temporal dynamics (interaction between variables or actors across time). For these types of information, Agent-Based Models (ABMs) are much better suited. To better represent food system complexity whilst incorporating the interactions among and within actors (businesses, consumers, etc.), there is a need for BNs and ABMs to interact dynamically.

Here, the first test of a fully-integrated model consisting of an ABM and of a BN is outlined, along with initial results from simulation runs. The integrated model uses data on consumers from the REFRESH pilot countries (from WP1).

3 Theoretical background

The structure of the “integrated ABM-BN” model is based on the Food Waste Model (hereafter FWM) developed within REFRESH WP1 (Van Geffen, van Herpen and van Trijp 2017). Thus, also the data for calibration of the integrated ABM-BN come from the WP1 questionnaire. The questionnaire, inspired by the Motivation, Ability and Opportunity theoretical model (Rothschild 1999; Thøgersen and Ølander 1995), tried to measure a set of fixed features and food-related behaviours and to quantify food waste within a sample of consumer households from the four REFRESH pilot countries (the Netherlands, Germany, Hungary, and Spain). The features detected can be grouped into six categories:

1. **Socio-demographics** (household size, age and sex of the respondent, income category, level of education, size of the municipality, and country);
2. **Motivations** (awareness of the consequences of food waste, attitudes towards wasting food, injunctive social norms, and descriptive social norms);
3. **Competing goals** with respect to food (health, taste, convenience as for preparing time, price, having enough food, and not having too much food at home);
4. Households’ **Food Waste Prevention (FWP) practices** (planning the shopping and using of food, avoiding buying impulsively, overviewing the food in stock, cooking precisely, and storing and using plate and pan leftovers);
5. **Opportunities** (availability of products, accessibility of stores, availability of space and storage equipment, and prevalence of unforeseen events);
6. **Abilities** (perceived difficulty with accurate planning, with creative cooking, with assessing food safety, and the knowledge of how to prolong shelf life);
7. **Psychographics** (awareness of parents, perceived financial control, and involvement in food preparation).

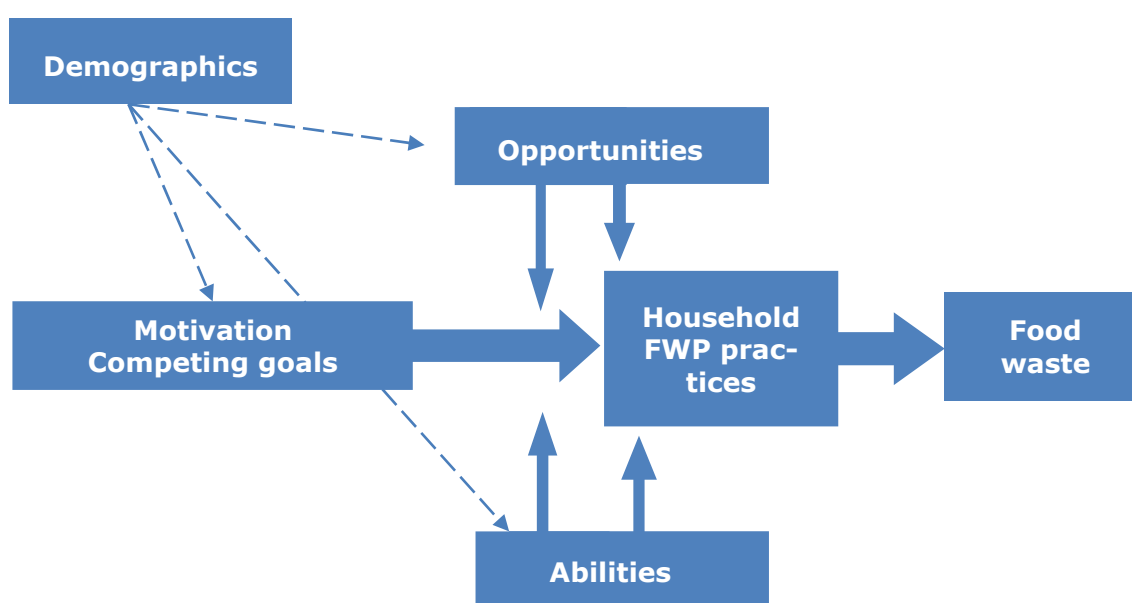
Figure 1 provides an overview of the relationships among these features according to the FWM elaborated by REFRESH WP1.

For the integrated ABM-BN model, the number of features considered (and their interrelations) were initially reduced to limit computational complexity, thus keeping computational time under an acceptable threshold, and to avoid cases where the BN could not generate a probability distribution of food waste due to the absence of agents presenting specific joint features (although this is still happening in some cases due to the small sample size of the questionnaire). However, the final model retains all features besides some of the socio-demographics: income, education and size of the municipality are excluded from the BN. Furthermore, in the ABM part, we introduce **salience**s as a measure of the importance, for a specific agent, of each competing goal, detected from a set of pairwise rankings included in the WP1 questionnaire. Finally, hereafter we refer to FWP practices as **actions**.

Within the WP1 questionnaire, all features apart from the socio-demographics were detected from the respondents’ answers to a statement (or the average of their answers to a number of statements) in a 7-point Likert scale measuring:

1. For the six **competing goals**, their importance (from “*not at all important*” to “*extremely important*”);
2. For the five **actions**, their frequency (from “*never*” to “*always*”);
3. For some of the **motivations** (awareness, injunctive social norms, and descriptive social norms), **opportunities**, **abilities** and **psychographics**, their level of agreement (from “*strongly disagree*” to “*strongly agree*”);
4. For the **attitude** towards for food waste, their approval (from “*strongly unfavourable*” to “*strongly favourable*”).

Figure 1: WP1 REFRESH Food Waste Model



When the level of a specific feature was measured by means of more than one item, the answers to all items were reported to a 1-7 scale, then the average was calculated, rounded, and transformed again into a qualitative answer using the levels listed above.

By construction, **saliences** range from five (none of the other motivations is more “salient” than the one at stake) to zero (all motivation are more important than it). Finally, the **socio-demographic features** are all categorical. A complete overview of all the elements included in the integrated ABM-BN model is provided in Table 1.

Table 1: Elements of the integrated ABM-BN model (variables).

Type	Variable name	Explanation / statement
Socio-demographic	country	Country of the respondent (Netherlands; Germany; Hungary; Spain).
Socio-demographic	sex	Gender of the respondent (dummy).
Socio-demographic	age	Age of the respondent (1 = ≤40; 2 = >40 & ≤56; 3 = >56).

Socio-demographic	household_size	Size of the household of the respondent (1 = one member; 4 = ≥ 4).
Motivation	awareness	Awareness of consequences of food waste (1 = strongly disagree; 7 = strongly agree) ¹ .
Motivation	attitude	Attitude towards wasting food (1 = strongly unfavourable; 7 = strongly favourable) ¹ .
Motivation	social_norm_inj	Injunctive social norms (1 = strongly disagree; 7 = strongly agree) ¹ .
Motivation	social_norm_descr	Descriptive social norms (idem) ¹ .
Competing goal	Health	Importance of food being healthy (1 = not important at all; 7 = extremely important).
Competing goal	Taste	Importance of food being tasty (idem).
Competing goal	Convenience	Importance of food being easy to prepare (idem).
Competing goal	Enough	Importance of having enough food at home (idem).
Competing goal	Price	Importance of food being cheap (idem).
Competing goal	Waste	Importance of not having too much food at home (idem).
Salience	S_health	Salience of health (1 = least important; 6 = most important).
Salience	S_taste	Salience of taste (idem).
Salience	S_convenience	Salience of time needed to prepare (idem).
Salience	S_enough	Salience of food security (idem).
Salience	S_price	Salience of price (idem).
Salience	S_waste	Salience of food waste (idem).
Action	action_planning	Planning of food shopping and use (1 = never; 7 = always) ¹ .
Action	action_impulsive	Impulsive buying (idem) ¹ .
Action	action_stocks	Overview of food in stock (idem) ¹ .
Action	action_precision	Cooking precisely (idem) ¹ .
Action	action_leftovers	Using leftovers (idem) ¹ .
Opportunity	opportunity_products	Availability of products (1 = strongly disagree; 7 = strongly agree) ¹ .
Opportunity	opportunity_store	Accessibility to stores (idem) ¹ .
Opportunity	opportunity_space	Equipment and space in home (idem) ¹ .
Opportunity	unforeseeable	Prevalence of unforeseen events (idem) ¹ .
Ability	ability_plan	Difficulty to plan accurately (idem) ¹ .
Ability	ability_cook	Difficulty to cook creatively (idem) ¹ .
Ability	ability_safety	Difficulty with assessing food safety (idem) ¹ .
Ability	ability_shelf	Shelf-life knowledge (idem) ¹ .
Psychographics	psychographics_parents	Parents' attention to preventing food waste (idem) ¹ .
Psychographics	psychographics_financial	Perceived financial situation (idem) ¹ .
Psychographics	psychographics_food	Food involvement scale (idem) ¹ .
Food waste	Totalgramsnoalc	Weekly household food waste in grams (from 0 to 7,075)

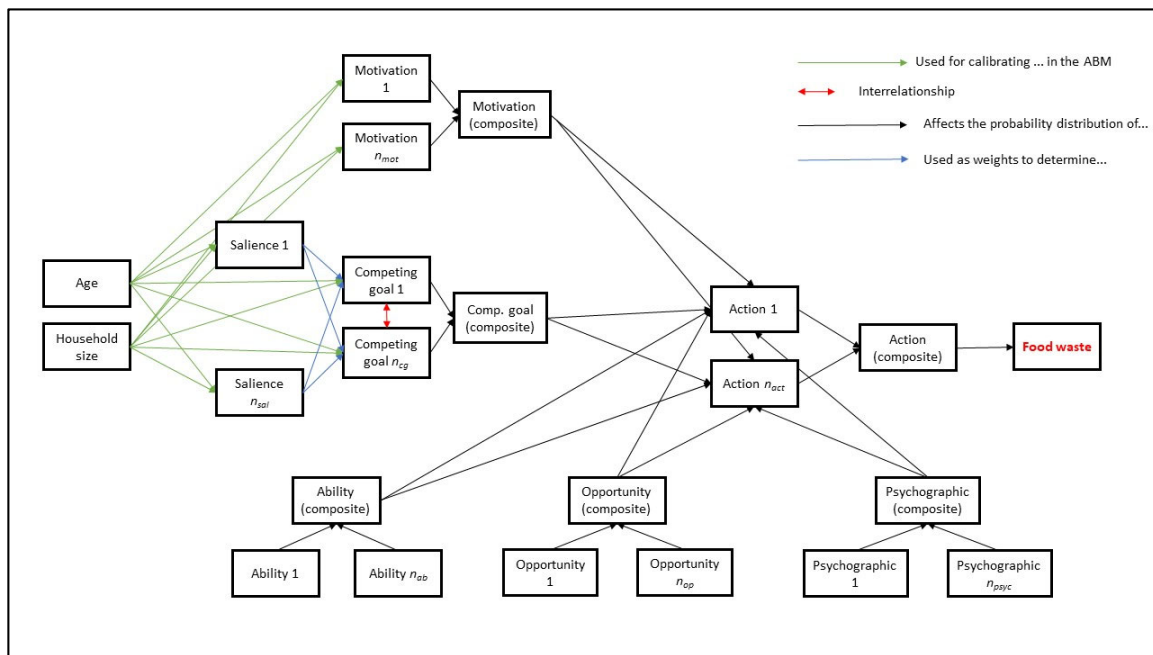
Notes: ¹ For the lists of the single statements of the questionnaire used to calculate the averages, see Van Geffen, van Herpen and van Trijp (2017), p. 91.

Source: Authors' elaboration on WP1 consumer survey.

4 The integrated model

The integrated ABM-BN model consists of a dynamic of opinions (represented by the ABM part), which are then turned into food waste levels through a set of conditional probabilities (the BN part). In turn, food waste levels influence the opinions of the agents¹ at the beginning of the following step. This process is repeated until food waste levels converge toward an equilibrium, or until the maximum number of steps set at the onset of each simulation is reached². Figure 2 presents a tentative structure of the interrelations among the nodes of the BN. It follows the theoretical structure of the FWM elaborated by Van Geffen, van Herpen and van Trijp (2017). However, the final structure of the BN, described later, is the result of machine learning, revised in order to approach the theoretical structure of the FWM. This compromise was motivated by the need of reducing computational time and by the fact that the machine learnt data did not return the structure of the FWM.

Figure 2: Scheme of the interrelationships within the integrated ABM-BN model.



Note: To avoid an excessive segmentation of the sample and preserve statistically significant distributions, the saliences, the competing goals and the motivations are calibrated based on the two socio-demographic features that influence food waste most according to the literature, i.e. age, and family size, rather than using all socio-demographics (see 5.1 Creation of simulated populations).

Source: Authors' elaboration on WP1 consumer survey.

¹ Agents are economic individuals (here, consumers) who act to achieve their purposes using heuristic or simple decision-making rules, and experience adaptation and interaction. Their behaviour is modelled in the agent-based model.

² In the preliminary simulations described in this report, the second strategy is used.

4.1 The Agent-Based Model

The ABM was developed in MatLab R2017a. Technical details are provided in the Appendix. After the simulated populations are generated with a process based on data from WP1, the ABM evolves according to the following intra-step dynamics:

1. For each agent i , one of the six competing goals m is selected for discussion;
2. For each agent i , the agents within her individual social network whose average opinion on the six competing goals is closer than a given threshold (neighbours) are selected for discussion;
3. The agent selected changes her opinion on competing goal m by averaging it with the average of her neighbours, with weights represented by the salience S_m and the complement to one of the salience $(1-S_m)$, respectively;
4. The opinions of the agent on all other competing goals change accordingly, based on an internal pulling mechanism;
5. The agent selected changes her *awareness* of food waste impact by averaging it with the average of her neighbours, with weights represented by the complement to one of her influenceability $(1-inf_i)$,³ and her influenceability inf_i , respectively;
6. The agent selected changes her *attitude* towards food waste by averaging it with the average of her neighbours, with weights represented by the complement to one of her influenceability $(1-inf_i)$, and her influenceability inf_i , respectively;
7. To measure injunctive social norms, the average attitude towards food waste of each agent i 's neighbours is calculated, with weights represented by each neighbour's salience of the competing goal "Waste"⁴;
8. To measure descriptive social norms, the median food waste of agent i 's neighbours at the end of the step $t-1$ (net of an error, due to the lack of visibility) is calculated, and then reported to a 1-to-7 scale, based on the position of the median in the food waste distribution of the population (septiles);⁵
9. If there were no neighbours within the threshold (point 2), thus no changes have taken place, the opinions of the agent on all motivations get back to their past values according to a "relaxation mechanism".

For the purpose of running the ABM, all levels assumed by motivations and competing goals are preliminary transformed from qualitative to a 7-point scale. Once the intra-steps described above have been completed for all agents, the new values of

³ One's influenceability $inf_i \in [0; 1]$ is set at the beginning of each simulation, with 0 indicating agents very difficult to influence, and 1 those very easy to influence. In the baseline, $inf_i = 0.5$ for all agents.

⁴ Then, the complement to 8 of the resulting average is computed, because in the statement used to measure the attitude towards food waste, high numbers represent a favourable attitude and *vice versa*, while injunctive social norms are measured as the level of agreements with statements indicating disapproval by close people.

⁵ We hypothesize that, if close people waste little, this can be equated to disagreement for the statement that one's close people throw away food regularly, and *vice versa*.

the six competing goals and the four motivations for all agents are sent to the BN⁶. The BN returns, for every agent, the probability that her individual food waste level falls within each of five classes⁷. Then, for each agent, a value of food waste is extracted from her individual probability distribution, assuming a uniform probability distribution within all classes and, for the last class, an exponential distribution with $\lambda = 3$ and $P(FW < 7,075) = 0.999$ (to reduce the probability of extreme outliers). Table 2 provides an overview of the food waste classes. After a food waste level is computed for every agent, a new time step of the ABM starts, in which the new food waste level is used as a parameter.

Table 2: Food waste classes for the extraction of individual values.

Classes (grams)	0	1 to 160	161 to 525	526 to 900	901 to 7,075
Original size	746	221	470	479	473
Original share	22.2%	20.9%	28.3%	14.5%	14.1%

4.2 The Bayesian Network

To integrate the consumer BN and the consumer ABM (Model 1), the first approach identified in D4.2 “Model development and data protocol” was followed. Real-world data (i.e., those of the four REFRESH pilot countries) was expressed probabilistically in a BN (technical details in the Annex). At each time-step of the model, ABM simulations of behavioural manipulations provide an input into a consumer BN⁸, which processes the effect of these on the distribution of food waste. The integration between AMB and BN allows the study of the effects of behavioural factors and social interactions on the evolution of individual opinions and actions regarding food waste (see D4.1a “Consumers behavioural economic interrelationships and typologies”). From the BN standpoint, consumer behaviour was operationalised as an “intervention” enacting upon the probability of producing one of five levels of food waste

⁶ In a previous version of the model, also the values of all “fixed features” (opportunities and abilities) were sent to the BN at each time step, although they were not changing at each time step. However, due to the very high requirements in terms of time and computational capacity, and to the fact that the model is not fully factorial (i.e. it was not possible to compute a food waste probability distribution for each and every combination of the conditioning variables, so that the model was returning zeros in many cases), it was chosen to set only competing goals and motivations. In future extensions, a limited number of “fixed features” at a time can be included in the ABM mechanism, thus being allowed to change at each step to simulate the impact of incentives or other policies.

⁷ In a previous version of the model, the 7-iles of the distribution of food waste within WP1 data were used. However, given the left-skewness of the food waste distribution, the model proved not so precise in predicting the average levels of food waste; thus, the classes were reduced to five to avoid unnecessary complexity.

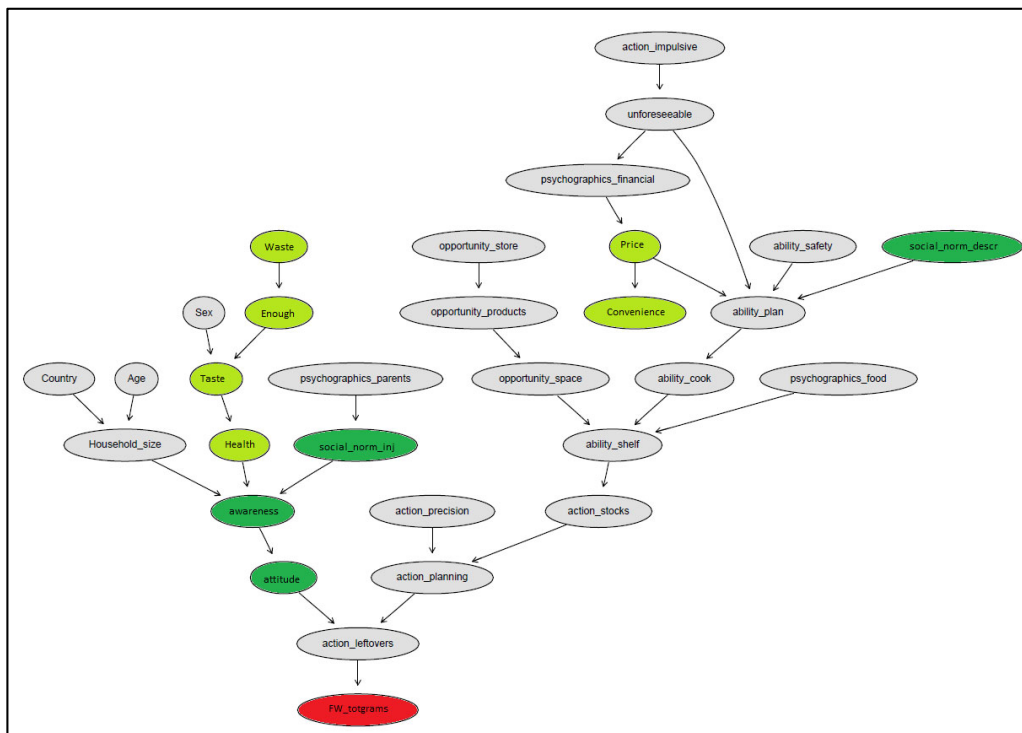
⁸ The BN was estimated starting from an Excel dataset obtained from the REFRESH WP1 survey, which included 31 categorical variables (those of Table 1, with the exception of saliences) and 3,354 observations from four countries.

(low to high) and by extending the cross-sectional data available to a dynamic setting. This allowed us to simulate the influence of human behaviour and interactions on the consumer food waste system.

The consumer BN was machine-learned, in order to identify the inherent structure of the data. Then, the arcs were reversed to obtain a BN structure compatible with the theoretical framework elaborated by Van Geffen, van Herpen and van Trijp (2017). This semi-structured BN represents a compromise between a fully structured model and a fully machine-learned one and is illustrated in Figure 3. The procedure of reversing the arcs to approach the structure of the FWM allows to partially overcome the issue that correlation does not mean causation. Indeed, while the arcs of a BN do not have a direction but only detect the presence of correlation, the causes and the effects and thus the direction were identified by relying on the theoretical FWM.

While the values of motivations and competing goals are set for each agent at each step, thus conditioning the value of the food waste node, the values of fixed features (opportunities, abilities, psychographics and socio-demographics) are used to estimate the BN, but no hypothesis on their value is made during the single time steps.

Figure 3: Graphical representation of the BN used in the integrated ABM-BN model.



Note: The nodes indicating motivations are in dark green, those indicating competing goals are in light green; their value (BN input) is set as a result of the ABM simulation. The food waste node (BN output) is in red.

4.3 The integrated model

BNs were developed in the open source statistical software R. The integration of the models was achieved through C++ in DOS, with externally controlled processes in both R and MatLab to allow the sharing of inputs and outputs.

5 Calibration and basic interventions

5.1 Creation of simulated populations

Simulated populations were created in MatLab using information from the WP1 questionnaire. First, the WP1 sample was segmented into 12 groups based on two variables identified in the literature as significantly related to food waste: household size (four categories, from one member to four or more members) and the respondent's age (three categories corresponding to terciles, i.e. less than 40 years, less than 56, and 56 or more). Thus, the first group was represented by households of one member aged less than 40, the second one by households of one member aged between 40 and 56, ..., the twelfth by households of four or more members whose respondent was older than 56. The incidence of each group in the WP1 sample, shown in Table 3, was calculated in order to replicate it in all simulated populations.

Table 3: Composition of the sample in terms of family size and ages.

Consumer groups	Absolute frequency	Relative frequency (%)	Cumulated frequency
1. One member, under 40	197	5.87	5.87
2. One member, 40 to 56	478	14.25	20.13
3. One member, over 56	387	11.54	31.66
4. Two members, under 40	213	6.35	38.01
5. Two members, 40 to 56	425	12.67	50.69
6. Two members, over 56	356	10.61	61.30
7. Three members, under 40	177	5.28	66.58
8. Three members, 40 to 56	329	9.81	76.39
9. Three members, over 56	61	1.82	78.21
10. Four or more members, under 40	219	6.53	84.73
11. Four or more members, 40 to 56	485	14.46	99.19
12. Four or more members, over 56	27	0.81	100.00

Source: WP1 consumer survey.

For each group, the probability mass functions of all competing goals, all saliences, and of the two motivations "awareness" and "attitude" was calculated, thus obtaining a set of matrices with as many columns as the groups (12), and as many rows as the states of the abovementioned variables (6 for the saliences, 7 for the others). These matrices, as well as a table with the relative size of the 12 consumer groups, were uploaded in MatLab. Then, after choosing the size of the populations simulated (1,000 agents in most cases), each agent was assigned to a group with probability equal to the incidence of this group in the WP1 sample. Finally, every consumer was assigned a value for each competing goal and salience, as well as for "attitude" and "awareness". Such values were extracted from the probability mass function of that specific feature for the group to which the agent belonged. While competing goals, "attitude" and "awareness" were expressed on a 1-to-7 scale (where 1 = "minimum" level and 7 = "maximum" level), saliences were turned into continuous values and

reported to a 0-to-1 scale, to be used as weights in the process of opinion averaging. Thanks to this procedure, every single agent in every simulated population was fully characterised in terms of goals, saliences, attitude, and awareness at time 1. The matrix for “awareness” is shown in Table 4 as an example. Besides that, all agents were assigned an initial value of food waste to be used to calculate descriptive social norms starting from step 2. This was obtained using a different procedure from the one just described for the other variables: for all of the 12 consumer groups, key quantiles of the waste distribution (10th, 25th, 50th, 75th, 90th, 95th and 99th) were calculated; then, a value included within these quantiles was extracted with probabilities equal to the difference of the quantile orders.

Table 4: Sample of matrix uploaded on MatLab (“awareness” of food waste effects).

Groups	1	2	3	4	5	6	7	8	9	10	11	12
1 (lowest)	0.7	1.7	0.8	1.0	0.8	0.0	0.0	0.5	0.0	0.6	0.0	0.0
2	2.6	3.7	2.6	1.4	4.2	3.4	2.8	0.9	1.0	0.9	1.7	1.7
3	4.1	10.1	7.3	3.4	7.6	6.6	5.2	5.4	5.2	6.5	8.1	1.7
4	26.6	25.5	24.3	26.5	24.2	23.2	21.3	24.9	19.6	22.2	22.5	10.2
5	34.7	32.9	30.2	33.0	26.5	31.7	31.7	25.3	21.7	33.5	26.2	22.0
6	25.5	18.5	24.5	23.4	27.3	26.0	26.5	28.1	38.1	26.5	26.2	37.3
7 (highest)	5.9	7.7	10.1	11.3	9.5	9.1	12.5	14.9	14.4	9.9	15.3	27.1

Notes: Cells report percentages for each demographic group, which sum to 100% per each column.

Source: WP1 consumer survey.

Finally, regression coefficients of each conflicting goal on all the other goals, and the associated standard errors, were calculated. These were included in two square matrices (the one with coefficients is alike Table 5), and uploaded in MatLab to implement the pull mechanism among conflicting goals – if a goal changes due to opinion exchange, the other five goals change accordingly, with the variation proportional to the correlation index plus a standard error. These tables represented all information needed to calibrate the ABM part of the integrated model.

Table 5: Regression coefficients between each motivation and all the others.

Motivation	Health	Taste	Convenience	Enough	Price	Waste
Health	1.000	0.232	0.048	0.185	-0.116	0.185
Taste	0.312	1.000	0.119	0.297	-0.082	0.023
Convenience	0.034	0.063	1.000	0.076	0.272	0.110
Enough	0.175	0.209	0.100	1.000	0.219	-0.229
Price	-0.071	-0.037	0.231	0.141	1.000	0.101
Waste	0.100	0.009	0.083	-0.131	0.090	1.000

Source: WP1 consumer survey.

5.2 Changes to baseline to assess impact on food waste

To assess the potential impact on food waste of interventions insisting on a specific element of the waste-generating mechanism, a number of changes can be applied

to the baseline populations. The variables to consider could be chosen based on their impact on the food waste node in the BN, e.g. if by changing their value it is possible to achieve a large percentage of entropy reduction in the food waste node (basically if, once set, they generate a conditional distribution of food waste with the smallest possible variance). Such a procedure implies that uncertainty in the state of the food waste node conditional on the “policy” node assuming a certain value is significantly lower than when no conditioning variable is set. The changes can be implemented one by one (single **policy**), or jointly (**policy mix**); then, by means of extensive simulations, it is possible to plot the evolution of food waste in the case of intervention and compare it to the baseline. Below, a number of potential interventions on different typologies of variables is outlined. Of course, these are non-realistic changes but allow us to “test” the responsiveness of the model to changes. In future model developments, evidence-based (if available) changes could be chosen to assess their potential impact on food waste levels.

1. For **opportunities**, setting the node “*opportunity_space*” at “*strongly agree*” for all agents equates to affirm that all consumers have enough space in their fridge, freezer and cupboard and enough kitchen supply (i.e. containers) to store their food and leftovers, which should result in less food waste. Such a change could be later linked to an innovation concerning fridges/freezers that, e.g., increases the internal space through thinner walls, or to a bonus to purchase larger fridges or freezer, which could be used only by households with enough kitchen space.
2. For **abilities**, setting the node “*ability_safety*” at “*strongly disagree*” for everyone implies that consumers are fully able to evaluate food safety and, thus, use food products and leftovers instead of throwing them away as a precaution. These changes could be promoted by means of incentives for retailers to provide training on these matters to their customers, or through the teaching food safety issues in school curricula.
3. For **competing goals** or **motivations**, setting the “*Waste*” goal at “*extremely important*”, the “*attitude*” towards food waste at “*strongly unfavourable*”, and the “*awareness*” of food waste consequences at “*strongly agree*” for all agents should result in less food waste. These changes could be achieved by means of an informational campaign focused on the negative effects of food waste for the society (e.g. environmental damage, waste of resources, inequality, etc.). Furthermore, a campaign that informs about food waste levels in the society, or about achievements in its reduction could be equated to an intervention on descriptive social norms where, besides their neighbours, consumers “discuss” with the legislator.

All the changes proposed can be implemented either at the onset of every simulation, or at a certain time step, including an evolution dynamic (e.g. through the **exchange of opinions** among agents, or through a rule for the **diffusion of innovations**). For example, consumers’ “*awareness*” of food waste consequences may increase either because they are hit directly by the informational campaign (they see videos, banners, etc.), or because they discuss with peers who have been hit by the campaign. Information about incentives to change the fridge can spread in the same way (thus networks make a difference), and potentially be limited to certain consumer categories. In the current version of the model, opportunities and abilities are **fixed features**; however, specific opportunities and/or abilities can be easily moved to the matrix of motivations and competing goals, thus being allowed

to change at each time step. Of course, since the BN model is not fully factorial (i.e. some combinations of values of the variables were not present in the dataset used to estimate the BN and, thus, it is not possible to calculate a probability distribution of food waste if this combination appears), increasing the number of conditioning variables means increasing the number of zeros (the consumers for whom the food waste distribution cannot be estimated) and thus the arbitrariness of the outcomes. Therefore, only a limited set of features can be subject to an intervention in a single simulation. A preliminary assessment of the impact of setting the levels of “*awareness*”, “*attitude*” and of the “*Waste*” goal respectively at 7, 7 and 1 is provided in the next section of the report.

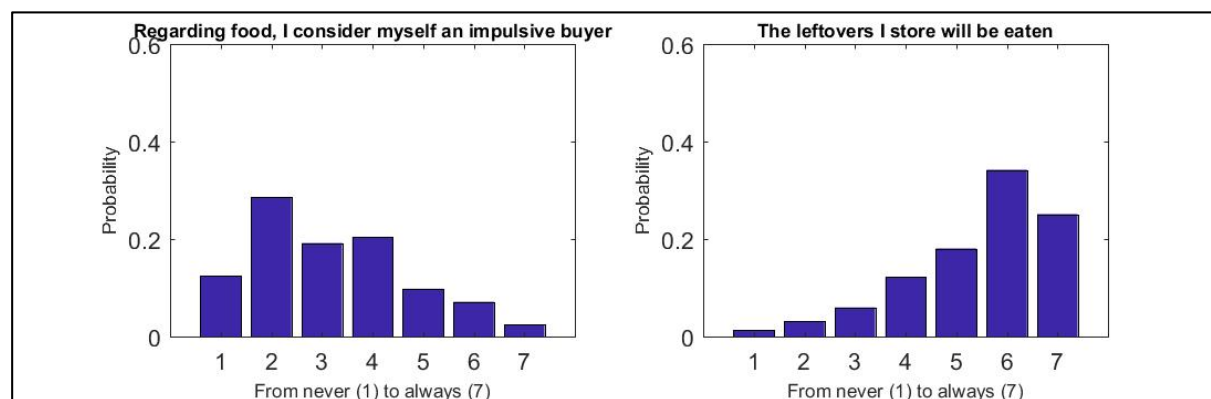
6 Results and discussion

6.1 Results of the simulations: baseline and changes

Before presenting the results of the simulations, it is worth illustrating the distribution of actions (i.e., household FWP practices) in the WP1 dataset. In the integrated ABM-BN model, the probability distributions of the actions within simulated populations are generated for each agent (at each time step) in the BN part by setting the values of motivations and competing goals for that agent (and for that time step). Then, the actions' distributions are mapped into a probability distribution of the levels of food waste.

Figures 4 to 7 provide an overview of the distributions of the self-reported frequency of household FWP practices before they are averaged to obtain the reduced number of features listed in Table 1. It emerges that distributions are **skewed**, with a peak at "usually" in most cases of right-skewness, and at "rarely" in most cases of left-skewness. Halfway answers prevail only for the frequency of shopping for food, and for impulsive buying. Only in the case of online shopping more than 50% of the sample answered "never". Instead, no clear trends are observed for actions such as measuring the ingredients of the meal, planning what to cook on a daily basis, and saving plate leftovers. These dynamics suggest that consumers tend to declare a high frequency for the actions perceived positively in the society, and a low frequency for those perceived negatively, pointing to the influence of a **social desirability bias**.

Figure 4: Self-declared frequency of actions (1-2).



The Figures from 8 on are based on series of 50-step simulations carried out on 100 populations. They illustrate the trend of key statistical indicators (plus or minus one standard deviation) along time. These simulated populations were created following the procedure described in Section 5.1, and differ only because of the process of random extraction of the values of motivations and competing goals (whose probability distributions are identical for all populations). Each population consists of 1,000 agents, and each simulation consists of 50 time-steps. In theory, a convergence criterion could be used to stop the simulation. However, it would not make sense at this stage, because no relevant policy measures which could cause a

change in its intrinsic trends are included in the model, and the nature of the relationships between variables (in particular the pull mechanisms between the conflicting goals) generates marginal persisting variations without convergence. Thus, a convergence criterion would have been very time demanding without improving our understanding of the system's dynamic. The average values assumed by relevant variables (food waste, motivations, and competing goals) were calculated for each population and for each time step, then obtaining means and standard deviations across populations. The level of food waste at time 1 was not plotted since, in theory, no food waste exists before agents define their motivations and goals and implement food-related actions.

Figure 5: Self-declared frequency of actions (3-10).

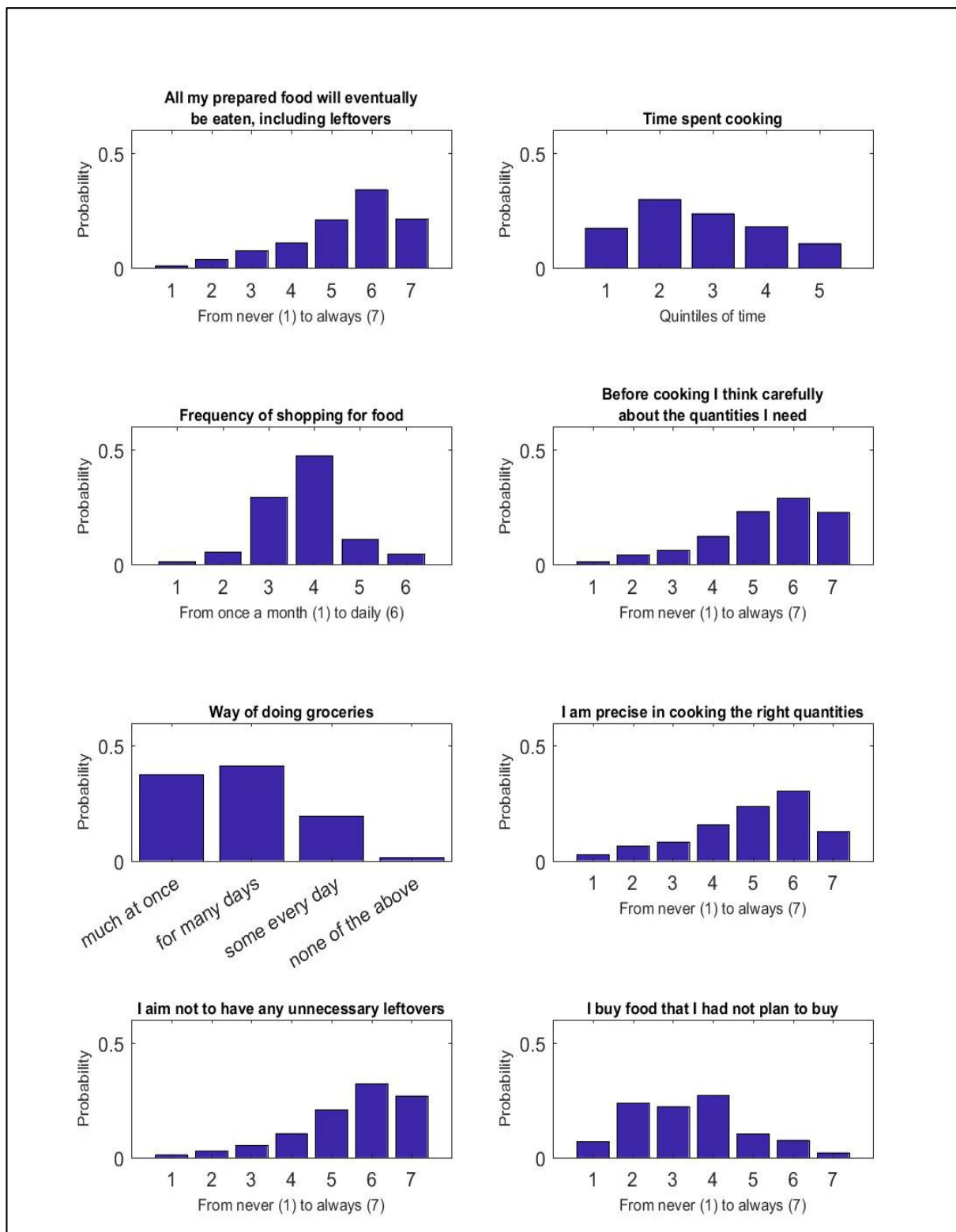


Figure 6: Self-declared frequency of actions (11-18).

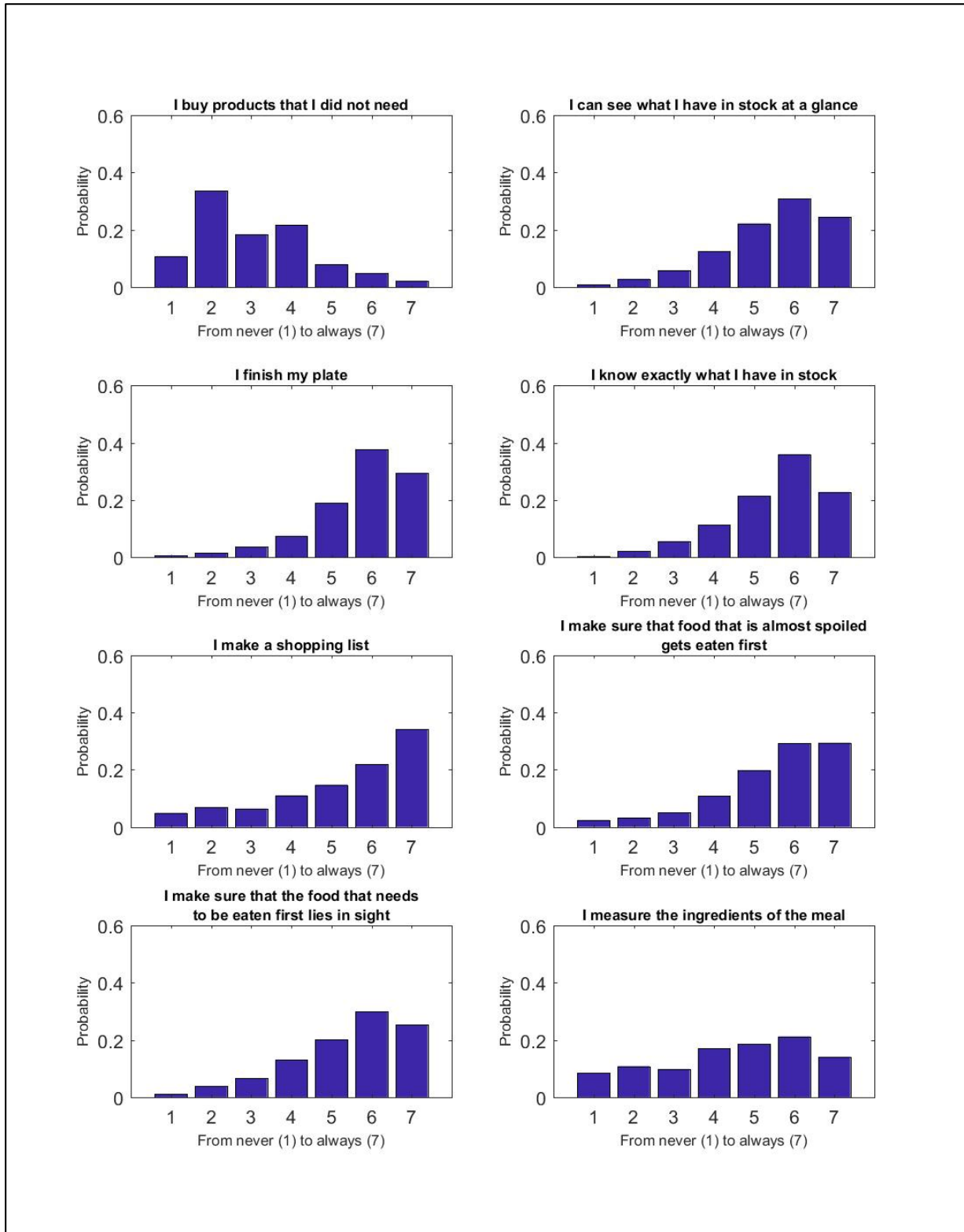
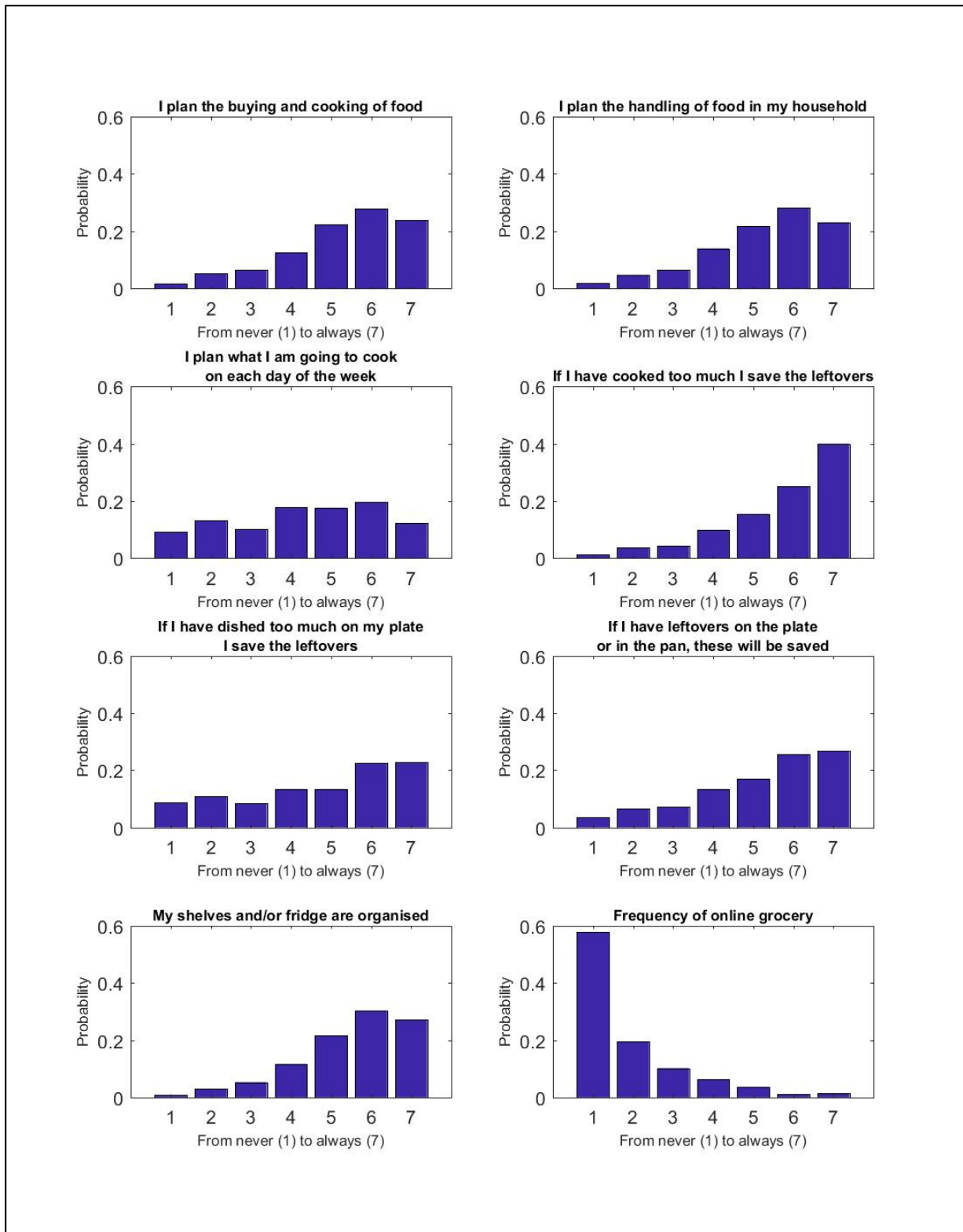


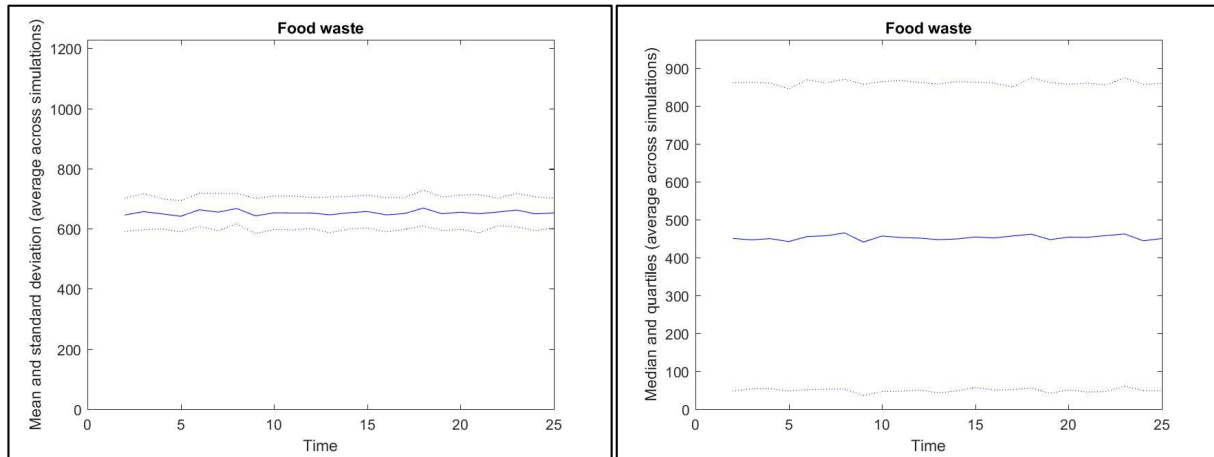
Figure 7: Self-declared frequency of actions (19-26).



The evolution of the mean (plus and minus a standard deviation) and of the median (and quartiles) of food waste along time in the baseline setting is shown in Figure 8. The average values of these synthetic indicators remain stable at around 650

grams for the mean and 450 grams for the median; regardless of the time step considered, a fourth of the agents waste less than 50 grams of food per week, another fourth more than 850 grams. The latter datum is particularly important, as these people should become the main target of any policies aimed at reducing food waste, while the lack of variation along time suggests that food-related habits are quite stable. The meaning of these indicators and their comparison with the values in the WP1 questionnaire are discussed in Section 6.2.

Figure 8: Food waste across time (mean and median) in the baseline setting.



However, Figure 9 shows that the percent of agents wasting less than 200 grams of food drops from around 45% to around 35% from the first to the last time step, while the percent of agents wasting over 2.0 kg per week becomes more relevant. This could mean that, if virtuous consumers are let free to exchange opinions on food-related goals and on food waste with careless people, they tend to be attracted towards average behaviours, i.e. to increase their food waste levels, probably because keeping them low requires commitment, which is difficult to sustain in absence of well-established habits. Moreover, opinion exchanges are probably increasing the visibility and, thus, the influence of positive outlier. Hence, communication can have both positive and negative effects. Overall, the standard deviation of food waste levels *within populations* is large compared to the mean (between 750 and 800 grams per week on average), and remains stable across time (Figure 10). Again, this is due to the influence of positive outliers while, at the onset of the simulations, a relevant share of consumers displays food waste levels close to zero and can thus be attracted only towards higher levels.

Figure 9: Food waste distribution at times 1 (WP1 data) and t_{max} (baseline).

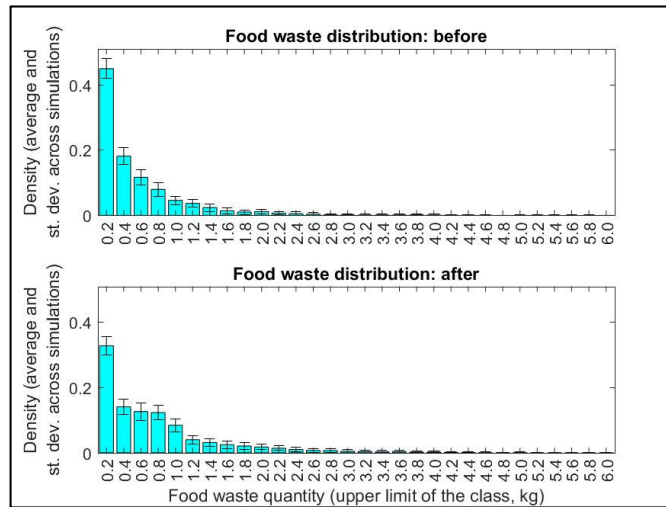


Figure 10: Standard deviation of mean waste across populations (baseline).

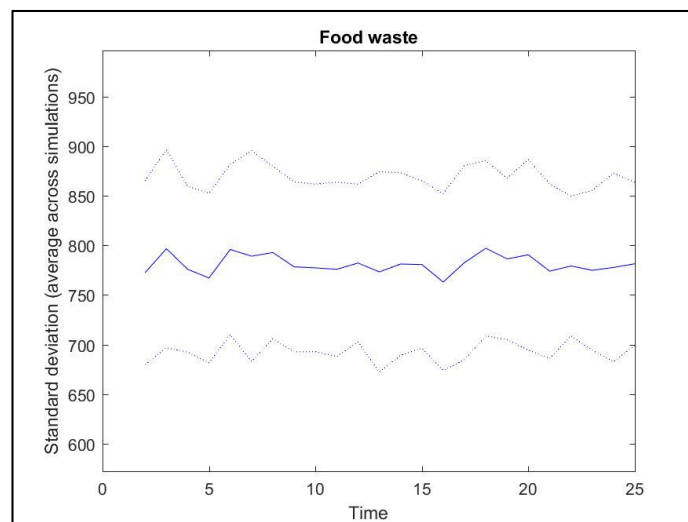
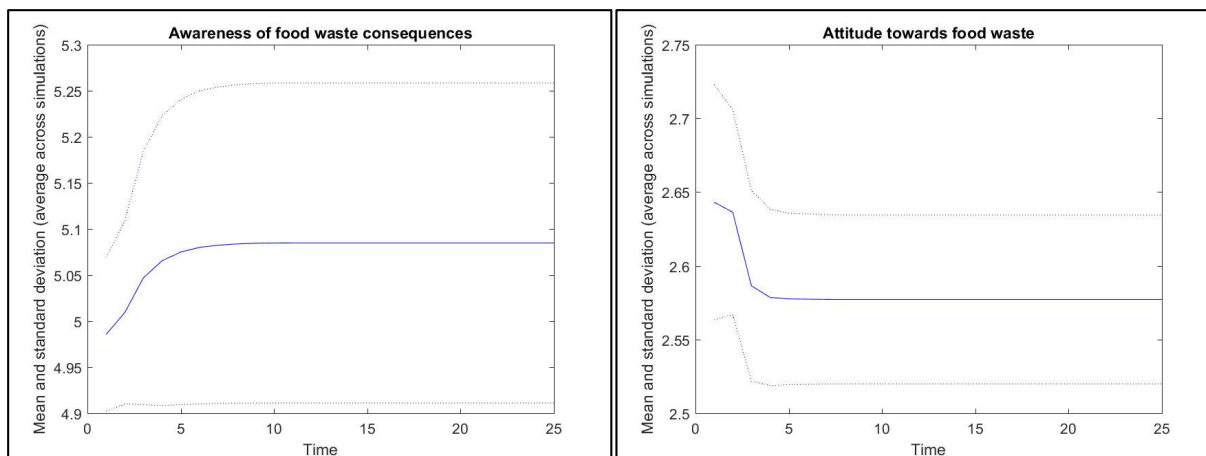


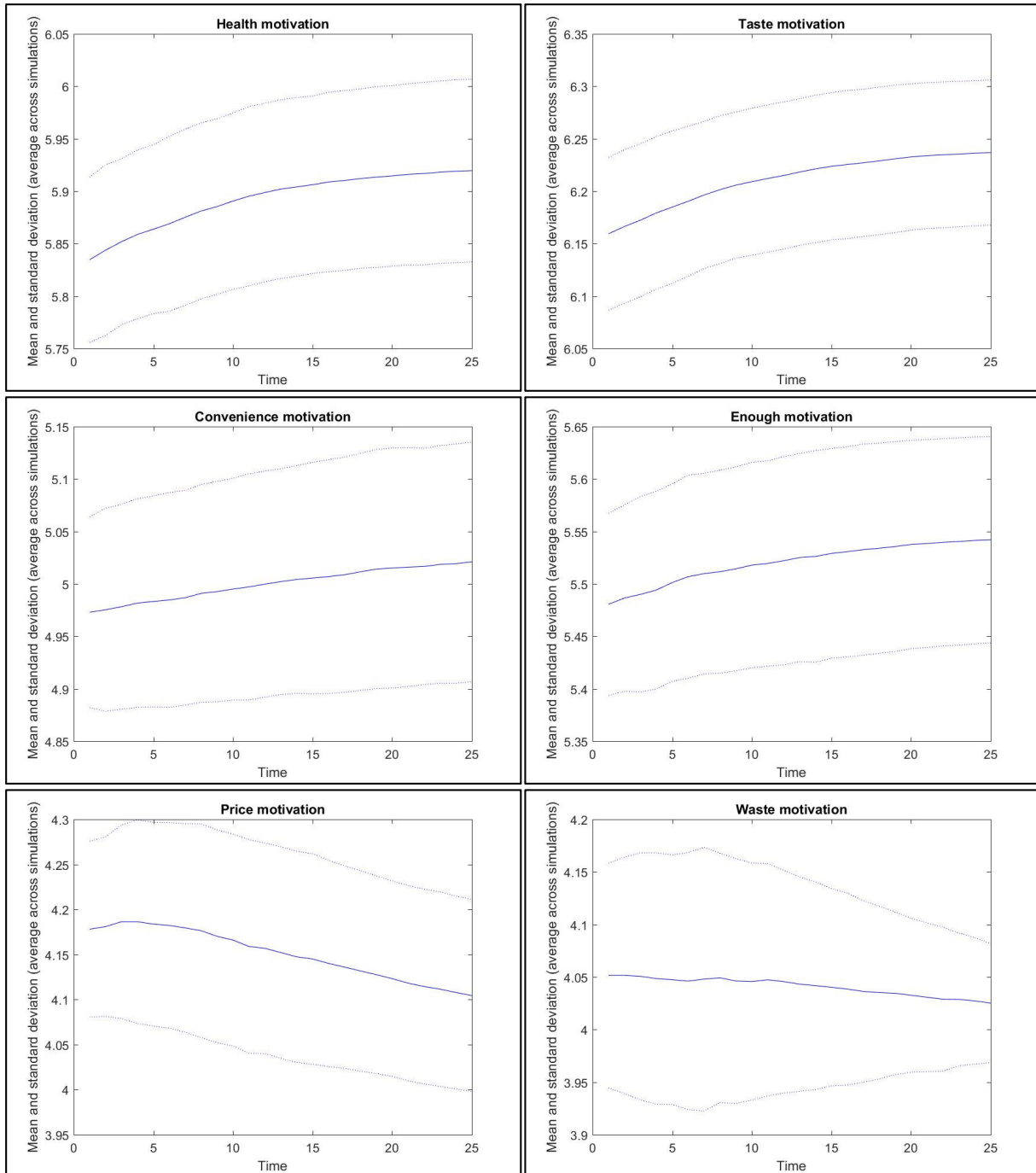
Figure 11 illustrates the evolution along time of the average *awareness* of food waste effects (where higher values indicate more awareness) and of the *attitude* towards wasting food (where higher values indicate higher acceptance), both measured on a 1-to-7 scale. Both motivations stabilise after the first five time-steps, because agents' individual networks, i.e. the people with whom every agent can discuss, do not change across time and no policy intervention is introduced to break up this equilibrium. This means that each agent's self-declared awareness (generally high) and attitude (generally low) tend to a linear combination between her initial levels and the average level of her neighbours, pointing to the role of opinion exchange in amplifying campaigns' effects.

Figure 11: Evolution of “awareness” and “attitude” along time (baseline).



Finally, Figure 12 illustrates the evolution of the average importance attached to the six competing food-related goals across time, measured on a 1-to-7 scale, where higher values indicate a higher importance. A trend can be identified for all of them: the importance attached to health, taste, convenience (in terms of time devoted to food management) and having enough food increases gradually, while the importance attached to price and non-wasting food decrease (moreover, the variability of the waste motivation across simulation narrows down with time). However, the absolute change is very small – around 0.05-0.10 points overall. This marginal trend is probably due to the pull mechanism: the goals positively correlated with most of the others tend to become more important, while those negatively correlated go in the opposite direction. Again, in absence of policy interventions that break up the trend, we observe a persistence of the dynamic inherent in the data or in the model design.

Figure 12: Evolution of competing goals related to food across time (baseline).



Now, a set of simulations with the levels of “*awareness*”, “*attitude*” and of the “*Waste*” goal set at a fixed value for all agents (respectively 7, 7 and 1) will be considered. The dynamics in term of food waste do not change much compared to the previous simulations, with stable values in terms of mean and median values. However, some counter-intuitive trends emerge, namely increased absolute values of food waste.

Figure 13: Food waste across time (mean and median) in the fixed values setting.

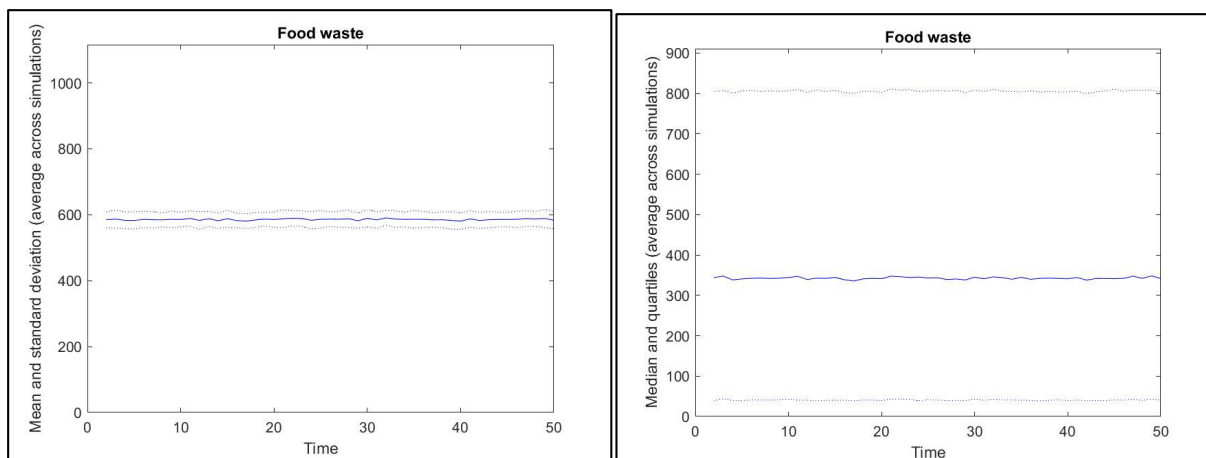
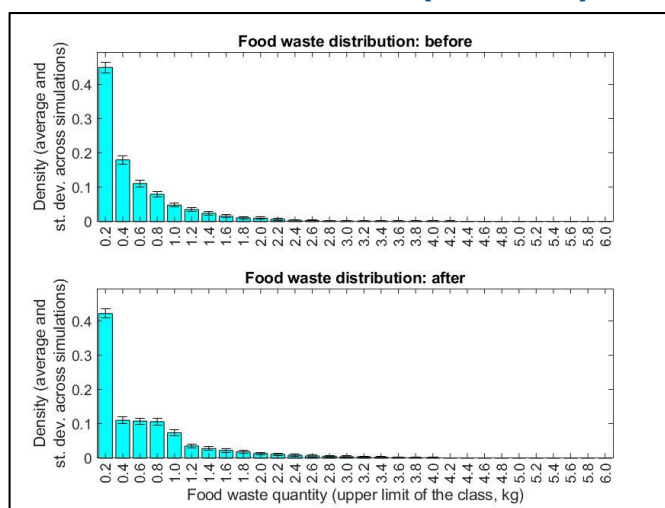


Figure 13 shows how the mean (plus and minus a standard deviation) and the median (and quartiles) of food waste evolve in a set of simulations with fixed values of “awareness”, “attitude” and of the “Waste” goal. The average levels of food waste remain stable around a value of less than 600 grams per week, considerably lower than the baseline setting. The median food waste levels across time are stable at around 350 grams, about 100 grams less than in the baseline setting. Even the quartile data are higher for this simulation: according to Figure 13, the lower fourths of the simulated populations produce almost 100 grams of food waste on average, while the higher 25% is set on values around 900 grams per week.

Figure 14: Food waste distribution at times 1 (WP1 data) and t_{max} (fixed values).

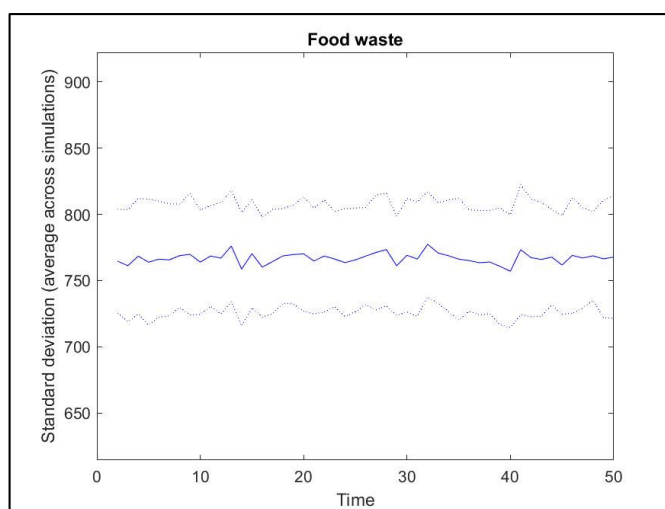


Concerning the distribution of food waste quantities at the beginning and at the end of the simulations (i.e. at steps 1 and 50, respectively), Figure 14 shows that the percentage of agents wasting less than 200 grams of food per week remains quite stable, dropping from about 45% at the first step to slightly more than 40% at the last step. The share of agents wasting 400 grams of food every week drops considerably, reaching a value of 10% at the last step of the simulation. These decreases are balanced by the increase in the average share of population wasting 800 grams of food or more, especially because the percentage of agents wasting more than 2.0 kg of food per week becomes relevant at the last time step.

Like in the baseline model, the average standard deviation of food waste levels within the simulated population is higher than the mean, and lays between 750 grams and 770-780 grams per week. These values are slightly lower than those recorded for the baseline models, and lead to the same interpretation on the role of outliers proposed in the baseline model.

In conclusion, a setting in which the values of “awareness”, “attitude” and of the “Waste” goal are all set at the level most favourable to food waste reduction, results in lower levels of food waste compared to a setting where those features are distributed in a more realistic way.

Figure 15: Standard deviation of mean waste across populations (fixed values).



6.2 Discussion of the results of preliminary simulations

This section discusses the fundamental features of the model, the main critical points and some potential improvements.

The model is essentially in equilibrium, with time-specific averages of the main variables oscillating around a central value derived from the data. This is as expected in the short-term, when the composition of the populations does not change, and in absence of either policies to reduce food waste or relevant shocks (e.g., food safety scandals, etc.). This proves the goodness of the integrated ABM-BN model. Indeed, there are no reasons to think that the populations of the pilot countries (from which the data for calibration are obtained) are not in equilibrium for what concerns food waste behaviours. However, it is worth pointing out that, starting from step 2 (i.e. excluding the initial step, when the conditions are based purely on WP1 data), the average weekly food waste per household in the simulated populations amounts to around 1.5 times its initial level (650 grams, compared to 441 grams). This happens because the probability of extracting a level in the highest class of the food waste distribution (900 to 7,075 grams) is higher than in the WP1 sample (14.1%) for a large number of agents. Indeed, when the BN cannot calculate a probability

distribution for an agent, the model extracts her individual food waste level assuming that the probability of being in a class amounts to 20% for each of the five classes⁹.

The evolution of the average motivations across time shows a certain trend, which is non-trivial even in a random network setting, and may signal the presence of a very gradual intrinsic change (e.g. because environmental concerns are spreading gradually). However, the variations are rather minor, demonstrating once again that the societies studied are substantially in equilibrium.

The effect of motivation changes on food waste is significant, while quite limited. This, we speculate, is due to three main reasons. First, in the WP1 questionnaire, food waste is reported as a quantity, rather than as a frequency. A strong motivation not to waste generates aversion to the act itself, thus reducing one's waste frequency; instead, the quantity of food wasted depends also on other aspects, which are not always under the control of the consumer (e.g. external temperature, products that were already damaged in the shop but whose conditions could not be assessed, etc.). This may cause committed households to waste more in quantitative terms than a family that cooks systematically more than needed, and throws away the leftovers. Second, in line with the theoretical framework adopted, opportunities and abilities moderate the relationship between motivations or competing goals and actions leading to food waste; thus, the impact of the former and the latter is not linear. Third, food waste quantities, as well as the frequency of the actions and the level of agreement with the statements measuring opportunities and abilities are self-reported; hence, they are subject to measurement errors or to underestimation/overestimation for social desirability bias. Finally, the model's output may simply reflect the underlying structure of WP1 data, where no clear correlations between "competing goals" and food waste are detected in most cases, and the regression coefficients obtained (as well as the R^2 summarising the goodness of fit) are very small (for instance, see the Annex of D1.4 "Quantified consumer insights on food waste").

The WP1 survey design as well as the structure of the integrated ABM-BN model have been influenced by the Motivation, Ability and Opportunity model as envisaged in Van Geffen, van Herpen and van Trijp (2017). This structure puts a strong prior on motivations driving food waste behaviour, but is improbable given the data collected in the WP1 survey (which identify a strongly significant but quantitatively marginal impact of social norms on food waste levels). As a robustness check, to assess the "importance" of all the variables, we built 1,000 Neural Networks from the full WP1 dataset using "Totalgramsnoalc" (i.e. the food wasted net of alcoholic products) as the output node. By summing the weights of the input variables, we can come up with a relative measure of their importance, reported in Table 6. This shows that what was disposed of, the age of the respondent, and actions related either to the organisation of the kitchen space or to leftovers, are all important in

⁹ Other testable options consist in assigning to the agents for whom no probability distribution of food was can be estimated: (1) their individual probability distribution in the last step in which this could be estimated; (2) their individual probability distribution at time $t = 1$, (3) the probability distribution in the population. The choice to assign the same probability to each food waste class was preferred since it implies no assumptions on the distribution and is, thus, the most neutral.

determining food waste. The ZIP code was the top-ranked variable, but probably this is an artefact of the large variability in the ZIP codes, rather than a geographic effect. Although this needs to be explored further.

Table 6: Relative “importance” of the variables in 1000 neural networks.

Variable	Importance (%)
ZIP code (not used in the integrated model)	100.000
Vegetables discarded (not used in the integrated model)	23.234
Meat discarded (not used in the integrated model)	22.756
Age of the respondent (Age)	12.314
Amount cooked	6.600
Soup discarded (not used in the integrated model)	5.058
Fish discarded (not used in the integrated model)	4.886
Legumes discarded (not used in the integrated model)	4.757
Topping discarded (not used in the integrated model)	3.336
Rice discarded (not used in the integrated model)	3.224
Taste motivation (Taste)	3.210
My shelves and/or fridge are organised (Act_stock4)	2.984
Egg discarded (not used in the integrated model)	2.932
Enough motivation (Enough)	2.818
Cheese discarded (not used in the integrated model)	2.765
If I have cooked too much, I save the leftovers (Act_leftovers3)	2.680
Size of the location (Municipality size)	2.467
Fruit discarded (not used in the integrated model)	2.427
Health motivation (Health)	2.421
Throwing away food gives me a feeling... (not in the integrated model)	2.376

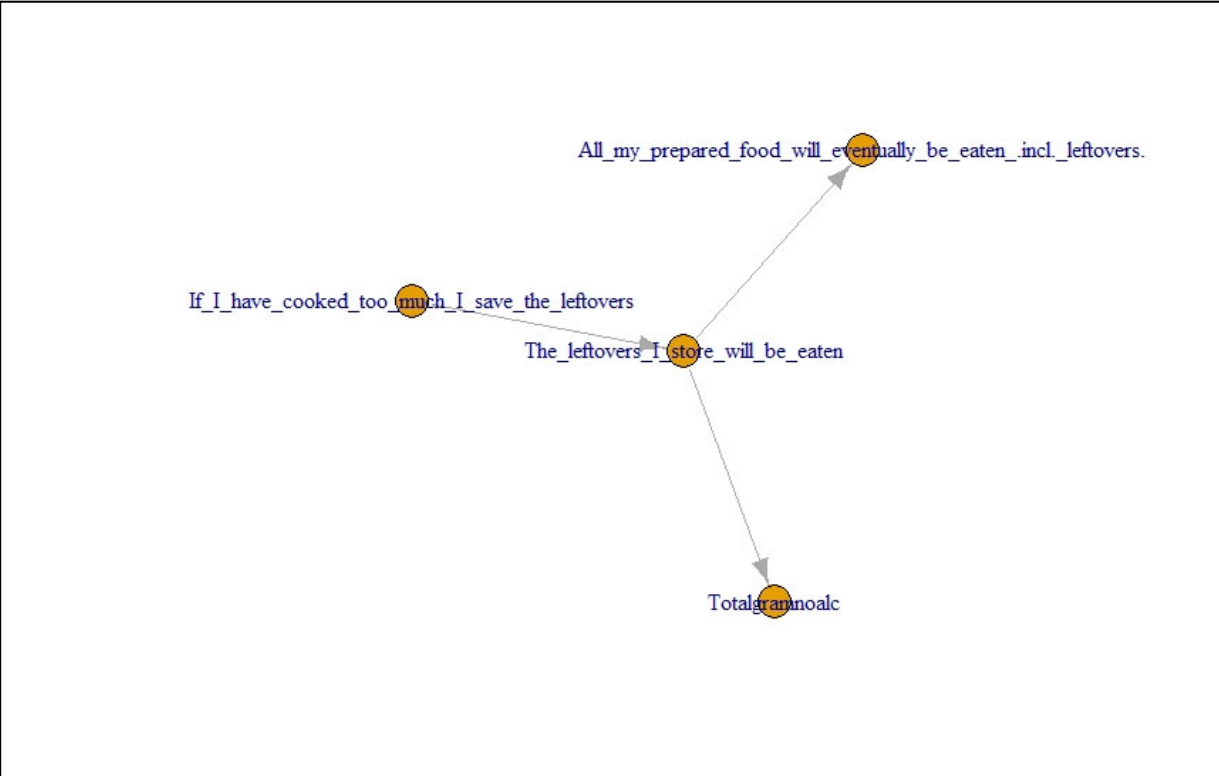
Source: Our elaboration on data from the WP1 consumer survey.

The final aim of this model is to obtain a tool which can be used to study the consumer food waste phenomenon at the national and European levels. To improve our results with a view to building a pan-European scenario of consumer food waste levels (the final goal of WP4 in the REFRESH project), in the BN we can reduce the number of nodes to those that are important (measured as a reduction in entropy). As a preliminary test, we did this using only the data on Dutch consumers (the most robust set of data from the REFRESH pilot countries, also due to the larger sample size). After building 1,000 Neural Networks and machine-learning the BN, we retained only the important and connected nodes, all relating to “*action_leftovers*”: “If I have cooked too much, I save the leftovers”, “The leftovers I store will be eaten”, “If I have dished too much on my plate I save the leftovers”). As a result, the predicted accuracy¹⁰ of the BN increased from an average of 47% to 69%. In

¹⁰ The coherence of the simulated results with the data from the WP1 survey.

the Netherlands, it seems that the amount of food waste produced by households is driven by the way people manage their leftovers. An image of the reduced BN estimated is provided in Figure 16. Using all country-level data together while inserting a country variable in the BN (to restrict the analysis to a single country when needed) can help avoid potential misinterpretations derived from the fact that data describing potentially different behavioural patterns are analysed aggregately.

Figure 16: Reduced machine-learnt BN of Dutch data from REFRESH WP1 survey.



Note: The direction of the arrows does not imply causality.

7 Conclusions and further applications

What are the benefits of the integrated model? The integrated model allows us to introduce the dynamic of interpersonal relations into the analysis of a policy intervention, while also taking advantage of the explanatory power of a Bayesian Network model to calculate conditional probabilities in the presence of a large number of variables. This allows us to simulate mid-to-long-term scenarios of consumers' food waste when only cross-sectional data are available. It also allows us to test the impact of a policy intervention beyond the people who directly experience the intervention. The model simulates the diffusion of information through opinion exchange, and its impact on actual consumer behaviours. The model is particularly suited to simulate policy interventions aimed at persuasion, such as informational campaigns, because of the Agent-based component, which is focused on social interactions between agents (consumers).

Further Research

The current version of the model is ready to assess the impact of certain policy interventions, such as campaigns, economic incentives, and trainings. However further research is required to understand the effect of an imposed structure on the model's predictive accuracy. Currently, the Bayesian Network relies on the structure imposed by the FWM developed in earlier REFRESH work (Van Geffen et al., 2017). Accordingly, the underlying conditional probabilities reflect the limited model fit of the data. The availability of larger datasets would allow to refine the model, relax some of the assumptions and therefore improve its predictive capacity. Our goal is to develop a roadmap for a 50% reduction in food waste by 2030. To that end, we will base the underlying structure of data on WP1 findings instead of imposing an existing one, and assess how that influences the distribution of food waste. We will also add Region variables (at NUTS2 level) into the model to account for country-level differences in behaviours, motivations and actions.

We also plan to allow some fixed features, namely consumer opportunities and abilities, to change as a result of ABM interactions. Since we plan to affect these features through policy interventions, we will add these to the matrix of changing features, and allow them to vary as a result of opinion exchanges, or of other diffusion mechanisms (Rogers [1962]'s model of innovation diffusion).

Preliminary results show that changing food-related motivations or competing goals has a limited impact on food waste. The strong prior belief on motivations implied in the "Motivation, Ability and Opportunity" model (represented by the model structure) was not supported by the structure of the integrated model, or by variable importance. We suggested a number of potential reasons for this that we need to explore further, as described in section 6.2.

What cannot be explained by the model, and why? Since the model builds on research of in-home food waste that does not include an analysis of different disposal or valorisation options, the integrated model cannot explain the impacts of how food waste is used after it leaves the household (e.g., animal feed, anaerobic digestion, etc.). It only investigates one phase of the global food waste chain, which is however one of the most relevant in terms of quantities generated. The model's focus is on internal household dynamics, where food waste represents a final output

and its destination is not relevant to later consumer decisions. Furthermore, since the Bayesian Network used in the model is data-driven, the outcome is influenced by the quality of data: if no correlation was detected between given variables (e.g. due to the small sample size, or because of biases of the respondents), no quantitative value can be assigned to the relation between them in the integrated model. Data on how motivations and food-related goals impact food waste is limited, and so the model's results reflect that (Van Geffen et al., 2017). This could be due to the inappropriateness of non-incentivised questionnaires in detecting consumers' true motivations and goals when an ethically sensitive issue like food waste is concerned. This is suggested by the skewed distribution of people's answers: they tend to answer in the way that people think society expects them to answer, potentially due to social desirability bias or cognitive dissonance. Thus, other research methods, like field or lab experiments, are needed to test the theory, elicit true preferences, and build a Bayesian Network that better resembles reality.

8 References

Godfray, H. Charles J., Beddington, John R., Crute, Ian R., Haddad, Lawrence, Lawrence, David, Muir, James F., Pretty, Jules, Robinson, Sherman, Thomas, Sandy M., and Toulmin, Camilla. 2010. "Food security: the challenge of feeding 9 billion people." *Science*: 1185383.

Grainger, M., Stewart, G., Piras, S., Righi, S., Setti, M., and Vittuari, M. 2016. *Model development and data protocol*, H2020 REFRESH, Newcastle-Upon-Tyne, UK: Newcastle University.

Gustafsson, J., Cederberg, C., Sonesson, U., and Emanuelsson, A. 2013. *The methodology of the FAO study: "Global Food Losses and Food Waste – extent, causes and prevention" – FAO, 2011*. SIK Institutet för livsmedel och bioteknik.

Thøgersen, J., and Ølander, C.F. 1995. "Perceptions of waste and recycling. A qualitative analysis." In *The 20th Annual Colloquium of IAREP, The International Association for Research in Economic Psychology, August 1995, Frontiers in Economic Psychology*. Bergen, Norway: Norwegian School of Economics and Business Administration.

Piras, S., Righi, S., Setti, M., and Vittuari, M. 2016. *Socio-economic implications of food waste: Business behavioural typologies and interrelationships*, Project report, EU Horizon 2020 REFRESH. Bologna, Italy: Bologna University.

Rogers, E.M. 2010. *Diffusion of innovations*. New York, NY, USA: Simon and Schuster.

Rothschild, M.L. 1999. "Carrots, sticks, and promises: A conceptual framework for the management of public health and social issue behaviors." *The Journal of Marketing*: 24-37.

Van Geffen, L., van Herpen, E., and van Trijp, H. 2017, *Quantified consumer insights on food waste: Pan-European research for quantified consumer food waste understanding*, Report of the EU project REFRESH, D1.4.

Xue, L., Liu, G., Parfitt J., Liu, X., Van Herpen E., Stenmarck Å., O'Connor, C., Östergren, K. & Cheng S. (2017) Missing Food, Missing Data? A Critical Review of Global Food Losses and Food Waste Data. *Environmental Science and Technology*, **20**: 51(12):6618-6633.

9 Annex

9.1 Technical details of the individual model components

9.1.1 Consumer Agent Based Models

MatLab runs a set of N_{Sim} simulations, each with a different population created based on the procedure described in Section 5.1. Every simulation consists of t_{max} steps. At every time step, N_{agent} agents are browsed. Each of them goes through a number of sub-steps, which result in new values for the six competing goals and for the four motivations; in turn, these are used by the BN to define a probability distribution of food waste levels for each agent. In the following pages, the features characterising the agents, their rules of interaction, and other relevant parameters are described.

Agent definition

Consider a set of N agents $i \in \{1, \dots, N\}$, who can represent either individual consumers or households. Every agent i is characterized by:

- A consumer **typology** T , based on two socio-demographic features found to be closely related to food waste in the literature (**age**, and **household size**), which constrains the values of her motivations, saliences and competing goals at $t = 1$;
- A set of six competing food-related **goals** $m \in M$ characterized by an opinion $o_i^m \in \{1; 2; 3; 4; 5; 6; 7\}$ (where 1 = "not important at all"; 7 = "extremely important"), whose initial value is constrained by the agent's typology T ;
- A value $s_i^m \in [0; 1]$ for each goal m , called "**salience**", that defines how relevant a goal is for that agent and, thus, how willing she is to consider others' opinions on that matter (0 = "not relevant at all"; 1 = "extremely relevant");
- A level of "**awareness**" of the effects of food waste $aw_i \in \{1; 2; 3; 4; 5; 6; 7\}$ (where 1 = "strongly disagree"; 7 = "strongly agree"), whose initial value is constrained by the agent's typology T ;
- An "**attitude**" towards food waste, $at_i \in \{1; 2; 3; 4; 5; 6; 7\}$ (where 1 = "strongly unfavourable"; 7 = "strongly favourable"), whose initial value is constrained by the agent's typology T ;
- An individual level of **food waste**, $FW_i > 0$, which is determined by the relationships of the BN after the values of motivations and competing goals are provided;
- An individual **network** F_i , represented by an adjacency matrix of size N_{agent} whose cells contain ones if the linkage between two agents exists, and zeros otherwise.

Even if every agent is characterized also by a value of **injunctive** and **descriptive** social norms, these are not an agent's characteristics properly, as they are a synthetic measure of, respectively, the attitude towards food waste and the food waste levels (net of an error) of the agents in her network.

Exchange of opinions

Agents are laid on a fixed social network. For the time being, this is an Erdős-Reyni random network with density λ .¹¹ Therefore, every agent has a social neighbourhood F_j that, in principle, can influence the values of her motivations and competing goals.

First, agent i selects, within her neighbourhood, the subset of agents $n \in N_i \subset F_j$ with $N_i = \left\{ j: j \in F_i \cap \frac{\sum_m s_i^m (o_i^{t-1} - o_j^{t-1})}{\sum_m s_i^m} < d_{int} \right\}$. This specification implies that agents are affected by some degree of “confirmation bias”, i.e. they are unwilling to compromise with individuals expressing opinions that are, on average, too far from their priors, especially if a goal is particularly salient for them. The “interaction threshold” $d_{int} \in [0; 7]$ can be defined as the distance beyond which another consumer’s opinion is not taken into consideration, thus measuring the inverse of the intensity of the “confirmation bias” expressed by a consumer.¹²

Once the agents $n \in Z \subset F_j$ are identified, agent i can potentially change her opinions on competing goals and motivations through a mechanism consisting of seven sub-steps.

First, one of the competing goals m is selected at random, and the opinion of agent i evolves as a result of discussion with her neighbours $j \in N_i$ according to the formula:

$$o_i^t = s_i o_i^{t-1} + (1 - s_i) \frac{\sum_{n \in N_i} o_n^{t-1}}{|N_i|}.$$

Basically, the agent gets nearer to the opinion-centre of the peers who are sufficiently close in terms of overall opinions. How much peers influence i is proportional to the complement to one of the salience that the goal m has for her. The logic behind this formula is that the more salient a goal is for agent i , the less willing she is to take into account what others *think* about it.

Second, if because of the exchange of opinions, agent i ’s opinion concerning the goal selected changes, all other opinions of her are reconsidered on the basis of an internal pull mechanism that makes use of a network of correlations whose signs and intensities assumed to be fixed at population level:

- the difference between the current and past opinion of agent i on m is computed $\Delta o_m^i = o_m^i(t) - o_m^i(t-1)$;
- for each goal $m' \neq m$, $\Delta o_{m'}^i = \Delta o_m^i (\beta_{m,m'} + r * se_{\beta_{m,m'}})$ is computed, where $\beta_{m,m'}$ is the regression coefficient linking m to m' , se_{β} is the standard error associated with this coefficient, and $r \sim N(0; 1)$;

¹¹ In the simulations illustrated in Section 6, the probability that a link exists has been modified and is equal, for agents i and j , to $\lambda * \left(1 - \frac{|\sum_m |o_i^m - o_j^m|}{6}\right)$ (i.e., proportional to the complement to one of their average distance in terms of competing goals, reported to a 0-to-1 scale).

¹² In the simulations illustrated in Section 6 $d_{int} = 7$, meaning that there is no confirmation bias.

- the new opinions of agent i on all goals $m' \neq m$ are computed as $o_{m'}^i(t) = o_{m'}^i(t-1) + \Delta o_{m'}^i$.

Third, agent i 's "awareness" of food waste effects evolves as a result of discussion with her close neighbours $j \in N_i$, according to the formula:

$$aw_i^t = (1 - inf_i)aw_i^{t-1} + inf_i \frac{\sum_{n \in N_i} aw_n^{t-1}}{|N_i|}.$$

Fourth, agent i 's "attitude" towards food waste evolves as a result of discussion with her close neighbours $j \in N_i$, according to a formula equivalent to the previous one:

$$at_i^t = (1 - inf_i)at_i^{t-1} + inf_i \frac{\sum_{n \in N_i} at_n^{t-1}}{|N_i|}.$$

Fifth, a time-specific value of injunctive social norms for agent i is calculated as the average attitude towards food waste of her neighbours, weighted by the salience of the competing goal "Waste" for the latter. This specification implies that neighbours for whom the goal of reducing food waste is more salient are more willing to spread their opinion and, thus, influential, and *vice versa*. The formula is the following:

$$inj_i^t = 8 - \frac{\sum_{n \in N_i} s_n^6 * aw_n^{t-1}}{\sum_{n \in N_i} s_n^6}.^{13}$$

Sixth, a value of descriptive social norms for agent i is calculated by first obtaining the median level of food waste of her neighbours at time $t-1$ (net of a noise), and then associating to it a value on a 7-point scale, based on its position in the distribution of food waste levels in the population at $t-1$. The formula is the following:

$$desc_i^t = K: \frac{K-1}{7} < F_{FW}(Md_{N_i}) < \frac{K}{7},$$

where Md_{N_i} is the median food waste level of agent i 's relevant neighbours (net of a noise), and $F_{FW_{t-1}}$ indicates the *cumulative distribution function* of this median in the distribution of food waste levels in the whole population. For example, if the median is located in the first septile of the population's distribution, consumers would have answered (in the WP1 survey) that they "strongly disagree" with the statement that their close people throw away food regularly, and so on.

Van Geffen, van Herpen and van Trijp (2017) found that one's food waste levels are not significantly affected by her beliefs about what "relevant others" *think* (injunctive social norms) but, rather, by her beliefs about what they *do*, i.e. their concrete food waste level (descriptive social norms). However, food waste is not visible: this is why agents' perception of others' waste levels is estimated with a degree of noise ε . The noise is normally distributed with average zero and standard deviation σ_ε .

Seventh, if $j \in N_i = \emptyset$ and, thus, agent i 's opinions on goals and motivations (here indicated as v_i^t) have not changed in time t , these tend to go back to their previous values, following a "relaxation mechanism" expressed by the formula:

$$v_i^t = v_i^{t-1} + \frac{(v_i^1 - v_i^{t-1})}{x},$$

¹³ Injunctive social norms are calculated as the complement to 8 of the average to reverse the 1-to-7 scale (see footnote 4).

where x is the “relaxation period”, i.e. the number of time steps required for goals and motivations to return to their initial value if they are not affected during the same number of time steps.

After applying the pull mechanism to all competing goals, as well as after calculating the new values of “awareness”, “attitude” and injunctive social norms, the new values are rounded to the nearest integer, to ensure that $\forall i$ and $v_i^t, v_i^t \in \{1; 2; 3; 4; 5; 6; 7\}$. This is done because the BN requires categorical variables, from 1 = “not important at all”/“strongly disagree”/“strongly unfavourable” to 7 = “extremely important”/“strongly agree”/“strongly favourable”.

Once the sub-steps described are accomplished, the new values of competing goals and motivations for all agents are sent to the BN. Based on this information, the BN returns the probability for each agent of her individual food waste being within each of the five classes shown in Table 2. Then, a value of food waste is extracted from the individual probability distribution of each agent, and a new step of the ABM starts.

Other relevant parameters

Besides calibrating the population on the basis of the procedure described in Section 5.1, at the onset of each series of simulations, other parameters are set:

- The number of simulations $N_{Sim} > 0$: for each simulation, a population is created – for the purpose of this report, $N_{Sim} = 100$;
- The maximum number of steps for every simulation $t_{max} > 0$ – set equal to 50;
- The total number of agents $N_{agents} > 1$ – set equal to 1,000 within this report;
- The network density (probability that a link exists) $\lambda \in [0; 1]$ – set equal to 0.1;
- The inverse of the intensity of the “confirmation bias” $d_{int} \in [0; 7]$ – for the purpose of this report $d_{int} = 7$;
- Agents’ influenceability when discussing food waste effects, to become aware of them, and attitudes towards food waste, $inf_i \in [0; 1]$ – set at 0.5 for every agent;
- The variance of the noise with which agents estimate others’ food waste $\sigma_\varepsilon > 0$ – set equal to 0.2;
- The “relaxation period” $x \geq 1$ – set equal to 10.

9.1.2 Consumer Bayesian Network Model

We used data from WP1 surveys in the REFRESH pilot countries to learn the conditional dependence structure between nodes. Model structure was determined from the FWM. For speed of processing we discretised the dataset (to convert continuous data to categorical data) and used this to build the models.

9.2 Data structure

Tables 7 and 8 contain respectively synthetic statistical indicators (average, standard deviation, median, minimum, and maximum) and the distribution of frequencies of all variables used for calibrating the integrated ABM-BN model.

Table 7: Synthetic statistical indicators for all calibration variables.

Variable	Average	St. dev.	Median	Min.	Max.
country ¹	2.440	1.212	2	1	4
sex ²	1.580	0.494	2	1	2
age ³	1.986	0.814	2	1	3
household_size	2.288	1.129	2	1	4
awareness	5.015	1.223	5	1	7
attitude	2.666	1.110	3	1	7
social_norm_inj	4.815	1.299	5	1	7
social_norm_descr	4.030	1.120	4	1	7
Health	5.875	1.170	6	1	7
Taste	6.163	1.022	6	1	7
Convenience	4.949	1.346	5	1	7
Enough	5.496	1.222	6	1	7
Price	4.177	1.460	4	1	7
Waste	4.072	1.504	4	1	7
S_health	4.778	1.386	5	1	6
S_taste	4.755	1.074	5	1	6
S_convenience	3.659	1.229	4	1	6
S_enough	3.651	1.205	4	1	6
S_price	2.065	1.261	2	1	6
S_waste	1.984	1.045	2	1	5
action_planning	5.119	1.214	5	1	7
action_impulsive	3.220	1.192	3	1	7
action_stocks	5.575	1.129	6	1	7
action_precision	5.179	1.226	5	1	7
action_leftovers	5.441	1.175	6	1	7
opportunity_products	5.474	0.930	6	1	7
opportunity_store	5.659	0.942	6	1	7
opportunity_space	5.637	1.076	6	1	7
unforeseeable	4.024	1.146	4	1	7
ability_plan	3.064	1.442	3	1	7
ability_cook	3.193	1.398	3	1	7
ability_safety	3.231	1.528	3	1	7
ability_shelf	5.355	1.111	6	1	7
psychographics_parents	5.589	1.243	6	1	7
psychographics_financial	3.846	1.689	4	1	7
psychographics_food	4.900	1.142	5	1	7
Totalgramsnoalc ⁴	2.773	1.325	3	1	7

Note: ¹ 1 = Netherlands; 2 = Germany; 3 = Hungary; 4 = Spain. ² 1 = male; 2 = female. ³ From this variable on, see Table 1 for the legend. ⁴ See classes defined in Table 2.

Source: WP1 consumer survey.

Table 8: Distribution of the answers for all calibration variables (%).

Variable	1	2	3	4	5	6	7
country ¹	30.68	25.07	13.83	30.41			
sex ²	42.01	57.99					
age ³	33.87	33.69	32.44				
household_size	31.66	29.64	16.91	21.80			
awareness	0.57	2.42	6.44	23.67	30.26	25.73	10.91
attitude	14.52	31.75	32.86	15.95	3.58	1.04	0.30
social_norm_inj	1.82	2.42	6.56	33.09	23.67	22.63	9.81
social_norm_descr	2.12	7.04	15.44	47.38	18.31	8.56	1.16
Health	0.89	0.98	1.25	9.06	17.50	34.97	35.33
Taste	0.81	0.63	0.75	3.34	12.61	37.48	44.39
Convenience	2.06	2.68	6.32	25.01	27.64	23.73	12.55
Enough	0.66	1.40	2.83	14.79	27.31	29.16	23.85
Price	5.25	8.05	13.36	34.20	21.88	10.76	6.50
Waste	7.60	8.11	11.90	36.46	19.44	10.76	5.72
S_health	2.74	6.89	8.71	14.67	25.49	41.50	
S_taste	0.51	3.31	9.15	20.51	40.25	26.27	
S_convenience	3.52	14.40	27.01	29.82	18.22	7.04	
S_enough	3.34	13.27	29.34	29.99	17.11	6.95	
S_price	43.41	28.89	14.04	7.30	4.14	2.21	
S_waste	40.79	32.35	16.85	7.72	2.30	0.00	
action_planning	0.15	1.91	8.53	18.28	29.28	30.20	11.66
action_impulsive	4.80	25.04	31.45	25.34	9.39	3.37	0.60
action_stocks	0.00	0.78	3.91	12.31	25.91	34.26	22.84
action_precision	0.33	1.82	7.39	17.77	29.04	30.02	13.63
action_leftovers	0.00	0.95	5.52	14.82	25.52	33.54	19.65
opportunity_products	0.33	0.18	1.73	12.05	30.14	46.42	9.15
opportunity_store	0.36	0.18	1.64	9.15	22.78	51.46	14.43
opportunity_space	0.03	1.01	2.92	11.09	21.23	43.65	20.07
unforeseeable	1.07	8.74	21.56	33.72	25.76	8.59	0.57
ability_plan	12.61	30.53	20.10	17.74	13.54	4.62	0.86
ability_cook	10.76	24.54	24.90	21.11	12.58	5.40	0.72
ability_safety	12.61	26.00	20.27	17.44	15.56	6.71	1.40
ability_shelf	0.36	1.55	3.91	13.63	29.52	39.03	12.02
psychographics_parents	0.69	1.43	4.32	11.36	21.56	35.30	25.34
psychographics_financial	9.18	18.10	13.36	21.44	19.14	13.80	4.98
psychographics_food	0.36	2.42	8.26	22.72	33.48	27.64	5.13
Totalgramsnoalc ⁴	22.24	20.90	28.29	14.46	14.10		

Note: 1 1 = Netherlands; 2 = Germany; 3 = Hungary; 4 = Spain. 2 1 = male; 2 = female. 3 From this variable on, see Table 1 for the legend. 4 See classes defined in Table 2.

Source: WP1 consumer survey.

9.3 MatLab and R codes

The MatLab and R codes used to run the integrated ABM-BN model can be provided upon request, addressable to any of the Authors of the report through the REFRESH website.