Model Description

The core agent-based model that we update and apply in this study was developed in Gibson et al. (2021) using the ODD (Overview, Design concepts, Details) protocol, a de-facto standard for documenting agent-based models (Grimm et al. 2006, 2010, 2020). Here, we focus on the aspects of model development and other specific elements of relevance, but provide additional details in Appendix 2. The framework for agent decision-making is given in Figure 1.

Overview

Purpose

The model's overall objective is to produce consumption trajectories from 2020 out to 2050 for dairy and plant-based milks (PBM). It does this by modelling individual-level preferences and food influences (informed by both theoretical grounding and empirical data) across physical, health and environmental perceptions, habit, social influence and the active evaluation of prior-choice. These future consumption curves (reported in average ml of milk per person per week) are directed, through parameter calibration via optimisation, to try and meet dairy reduction targets posed by UK bodies for 2030 and 2050. Specifically, the study performs simulation experiments to assess and compare six different milk consumption scenarios that are distinguished by differing model assumptions and target level.

Agents, state variables, scale

Agents represent consumers that each have a disposition to consider (or not) their milk consumption choices. Agents construct a cognitive choice function for dairy and PBM, comprised of the perceived physical (modelled as price), health and environmental characterises of each choice. These are computed at each time step of the simulation, and are modified by other food influences (habit, social influence) and choice evaluation, each governed by individual sub-models. The relative importance that agents ascribe to the physical, health, environmental, habit and social influences is determined by empirical data operationalised from the British Social Attitudes (BSA) 2008 survey (National Centre for Social Research 2010). The use of survey data to construct agent characteristics is a common approach in agent-based modelling. For example, Scalco et al. (2019) also use BSA data in their study of UK meat consumption, and Khademi et al. (2018) use the California Health Interview Survey to inform their ABM of health-eating in Los Angeles. Upon an agent becoming ‘disposed’, the quantity of each milk option is apportioned according to the relative size of each total choice function.

Agents have an existing choice mirroring the average consumption levels of dairy and plant-based milks in 2019, and form a social network. Social influence is stochastic, occurring as a function of interaction probability, modulated by the relative importance an agent places on social influence in food, derived from the BSA 2008 survey data.

Process overview and scheduling

At each model run, 1,000 agents are created (see Figures A3-A5 in Appendix 1 for an exploration of different agent population size and resampling of agent attributes from survey data), initialised with an incumbent choice, reflecting the consumption split between dairy milk and PBM in 2019, and randomly linked with other agents in a network (see social influence sub model). Agents construct a choice function based on perceived information about milk characteristics, employing memory effects to draw on information perceived in previous time steps. Other food influence factors; habit, social influence, and the evaluation of choice, all impact the final choice functions. The quantity of each type of milk consumed is calculated at each time-step (annually from 2020 to 2050), based on total choice function scores (see next section for details).

Design concepts

Basic principles

The basic structure of agent decision-making is given by the process flow diagram in Figure 1. In brief: agents perceive physical, health and environmental characteristics of each milk choice; agents are then triggered (or not) to enter a state of disposition to consider their milk options; a quantified choice function is calculated for each option, comprised of the perceived characteristics and modified by habit and social influence; agents may evaluate their choice and inform future decisions based on internal consistency between the impact of their choices and the human values they hold.

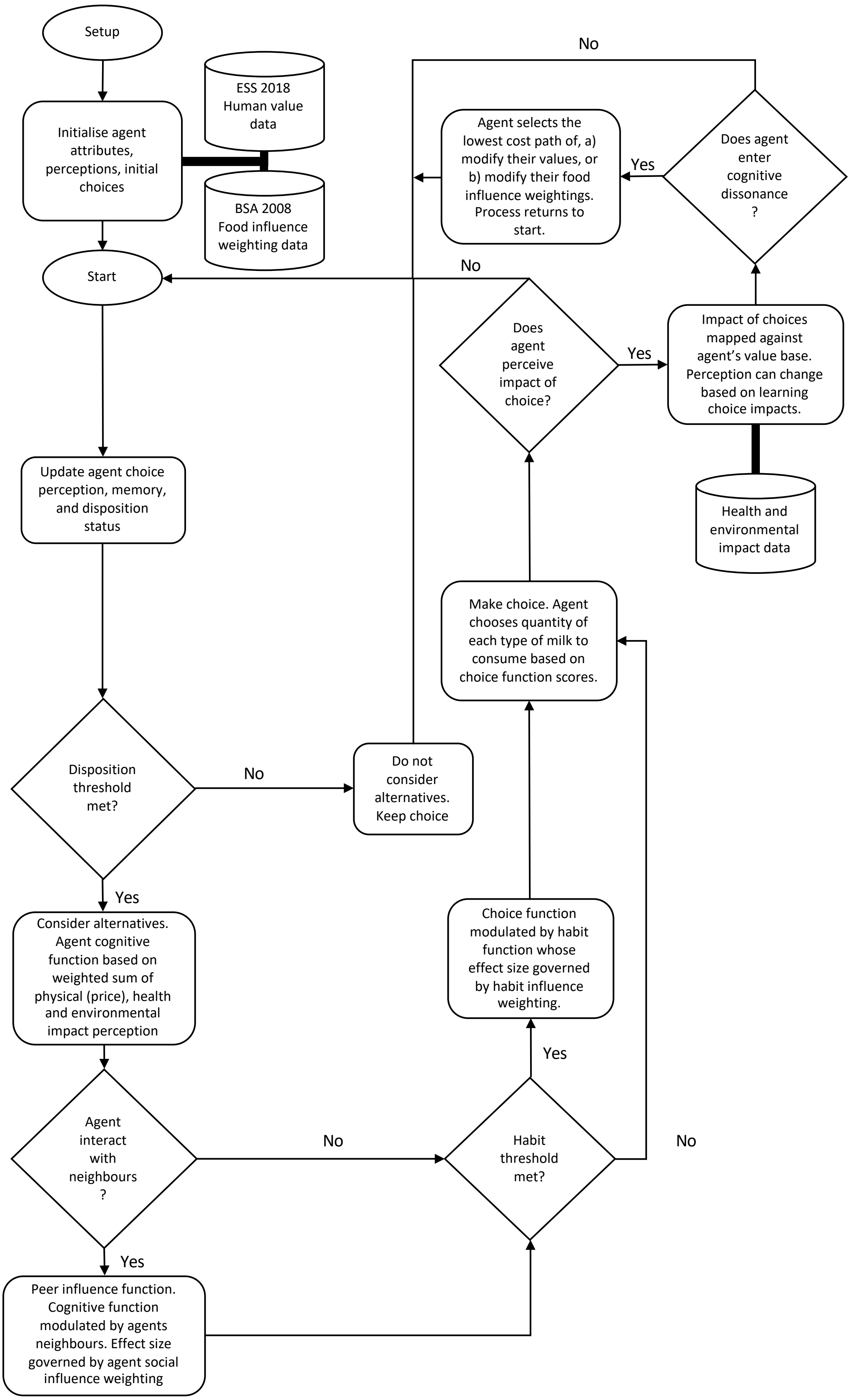

Figure 1: Flow diagram of agent milk choice influence and decision-making in the ABM.

The key functions that act upon agent choice are described in the sub-model section. Table 1 gives the parameters that operate within the model, a more detailed description of which is provided in Appendix 1 (Table A1).

Table 1: Model parameters, value ranges, and the sub-model in which they operate.

Parameter	Sub-model	Dynamic	Range
1. Memory length	Cognitive perception	No	[1,10]
2. Habit threshold	Habit	No	[1,10]
3. Probability of interacting	Social influence	No	[0,1]
4. Initial habit of incumbent	Habit	No	[0,10]
5. Social blindness	Evaluation	No	[0,1]
6. Post-choice justification	Evaluation	No	[0,1]
7. Cognitive dissonance threshold	Evaluation	No	[0,1]
8. No. of neighbours	Social influence	No	[2,10]
9. Perception of health impact of PBM	Cognitive perception	Yes	[1,3]
10. Perception of environmental impact of PBM	Cognitive perception	Yes	[1,3]
11. Gradient of probability disposition	Disposition	No	[14,16]
12. Perception of health impact of dairy	Cognitive perception	Yes	[1,3]
13. Perception of environmental impact of dairy	Cognitive perception	Yes	[1,3]

At each time-step, the milk consumption of agent's that are 'disposed' to consider their choices is given by the relative proportion of each option's choice function of the total summed choice function, multiplied by the total milk consumption. I.e. if dairy and PBM have the same choice function value, an agent will consume 50% of the total available milk (expressed as ml per person per week) for each choice. Agents that are not in a state of disposition at a given time-step repeat their previous choice and milk consumption.

In the model, total average weekly consumption was maintained at 2019 levels over the simulation period out to 2050. This was because we were primarily concerned with product substitution rather than absolute decrease in consumption. However, this is a clear motivation for future work, and as a starting point, Figure A2 in Appendix 1 shows a simple extension to the scenario analysis by considering future non-constant (declining) total consumption.

Details

Implementation and initialization

The model is implemented in NetLogo 6.0.4 (Wilensky 1999), a copy of which, along with associated Python code, is available at https://www.comses.net/codebase-release/1bcb23b1-92b3-4974-bdef-11abeed3e6d3/.

Input data

Agents are initialised with basic human values (Schwartz's universalism and security values (2003, 2006, 2012)) using data operationalised from UK specific responses (n=2,167) of the 2018 European Social Survey (ESS) (Norwegian Centre for Research Data 2018). The survey questions associated with these values are reproduced in Table A2 of Appendix 1. Survey responses were on a six-point scale of 'Very much like me' to 'Not like me at all' and also included an additional three coding options of 'Refusal', 'Don't know', and 'No answer'. Specifically, we take cross-tabulated data of the two relevant question responses, weighted to account for differences in selection probability from the sampling design, and obtain the proportions that cover each of the 36 possible response combinations. A random sample of these responses is taken, equal to the number of agents modelled. Here, this was typically 1,000, which was tested with resampling and different agent population size – see Appendix 1. The sample is loaded into the model and each agent is assigned a universalism and security attribute accordingly. A uniform probability distribution determines the specific number that this takes, with the six-point response scale converted into six equal sized bins between 0 and 1. E.g., an agent with a 'Very much like me' response will have an equal probability of scoring between 0.833 and 1.000, and so on.

Agents are also assigned a weighting for each of the main food influence categories in the model (physical, health, habit, social, environmental). These weightings are operationalised from British Social Attitude (BSA) 2008 survey data which included a series of questions on food influences. BSA 2008 contains 4,486 survey responses across a number of social attitude and demographic dimensions. This study was interested in the section on food influence, which contained 19 direct influences (and options for 'other', 'someone else decides' and 'no particular' influence). Responses were recorded as either 1 (having an influence), 0 (not having an influence), or -2 (did not answer). After removing null responses (-2 values), 2,238 responses remained, of which 1,000 were randomly sampled for inclusion in the model to directly represent agent influences (as with ESS, multiple samples were tested at 1,000 and different samples sizes were drawn - see Appendix 1. The 19 different influences were assigned to one of five most closely aligned categories (physical, habit, health, social, environmental), and then each category was scaled, so that categories had the same total representation, summed and converted into a proportion of the total summed influence. The mean weights across the sample were: physical = 0.344, habit = 0.214, health = 0.204, social = 0.123, environmental = 0.116. See Table A3 in Appendix 1 for details of the BSA survey questions.

Sub-models

Disposition

Gibson et al. (2021) compare two mechanisms of disposition, a threshold-based, and a probability-based approach. From that study, it was found that the latter performed better in reproducing observed macro level data of historic milk consumption. And so, here, we opt to employ the same probability-based disposition approach, which itself was influenced by previous studies modelling agent disposition dynamics in social networks (Galán et al. 2009; Wang et al. 2017). The probability to become disposed to consider milk choice options is based on how alike an agent's neighbour choices are, and uses information entropy to calculate maximum and minimum 'alikeness'. Equation 1 expresses this disposition function:

$$p(disposition) = \frac{1}{1 + \exp\left(-k \cdot \left(\frac{h}{h_{\max}} - 0.5\right)\right)}, \qquad (1)$$

where k (parameter 11 in Table 1) is the gradient of the probability logistic function; 0.5 is a coefficient to limit values between 0 and 1; and h/h_{max} gives a proportion of how homogenous or heterogenous an agent's neighbours aggregate choice is (see Equation 2), where h_{max} equals 1 ($-\log_2 0.5$). Equation 2 expresses neighbourhood choice information entropy:

$$h = -\left(\frac{f_{dairy}}{f_{all}}\right) \cdot log_2\left(\frac{f_{dairy}}{f_{all}}\right) + \left(-\left(\frac{f_{PBM}}{f_{all}}\right) \cdot log_2\left(\frac{f_{PBM}}{f_{all}}\right)\right), \qquad (2)$$

where, f_{dairy} and f_{PBM}, are the frequency of an agent's neighbours that choose dairy milk or PBM, and f_{all} is the total number of neighbours.

Cognitive perception

The cognitive perception sub-model represents how information regarding different milk choice characteristics are perceived by agents. Central to this are the calibrated health and environmental perception parameters, the value of which is varied to reflect its non-constant nature. I.e. perception of something can change with time, space, context, and of course different individuals. Values are drawn from a normal distribution, where the means of these distributions are determined by the perception parameters, with standard deviation of 0.1. A normal, rather than say a uniform, distribution is chosen as it gives a higher and symmetric probability of producing a value close to the calibrated mean, while still allowing the chance of values to deviate strongly from this.

For scenarios that consider current and organic pricing, means are taken directly from a fixed value representing the relative price relationship between dairy and PBM. Price data on PBM and organic dairy was collected online (in November 2021 via manual means) from publicly available data from three major UK supermarkets (Tesco, Sainsbury's, Morrisons). Average prices for conventional dairy milk were calculated from UK Family Food Survey 2017/18 data. Mean values were; 164p/l for PBM, 106p/l for organic dairy, and 60p/l for conventional dairy. Note, prices for PBM and organic milk were simple averages and not weighted by product volume sold.

These prices were operationalised so as to enable adequate inclusion in the cognitive function. Here, a larger price is a negative characteristic, and the model treats overall choice as a positive sum of all the different influences. The PBM price was set at 1 (most expensive, therefore lowest score) and the price multiplier between PBM and current conventional milk or organic milk was assigned to these options accordingly. This resulted in a value of 2.72 for current conventional milk, and 1.55 for organic milk. That is to say, PBM is 2.72 times more expensive than conventional milk, but this is represented as a positive 'bonus' for dairy. This may not be the optimal approach if we were concerned with more granular realism, but for the purposes of this study, this abstraction was deemed a reasonable proxy.

The cognitive choice function (Equation 3) is comprised of the three modelled milk characteristics, weighted by the relative importance placed on it (out of the five influence categories assessed).

$$f(cog.) = \beta_{phy} \cdot (physical) + \beta_{hel} \cdot (health) + \beta_{env} \cdot (environmental), \qquad (3)$$

where β_{phy}, β_{hel}, and β_{env} are the weights assigned to the perception of physical, health, and environmental aspects of the milk choices. At initialisation, agents are assigned a set of weights drawn from BSA 2008 survey data (see 'Input data' for more details).

Environmental concern

In this function, exogenous changes to agent environmental-based choice influence are modelled. It consists of two variables: a probability of occurrence, and a magnitude of effect. The two different approaches that scenarios S5 and S6 test are constructed from YouGov weekly/monthly public concern issue tracker data (YouGov 2021). Here, we approximate the longitudinal change in UK public environmental concern (given as a % of people that rank 'environment' as a top issue) as the size of potential percentage change in environmental weighting of milk choice influence (β_{env} from Equation 3). This percentage change in weighting is added to an agent's existing environmental weight, and subtracted from its physical (price) weight (β_{phy}). This ensures that the total influence weighting remains equal to 1, with the model controlling for any weight values that would be outside of the 0 to 1 range.

In the case of scenario S5, the probability of a shock occurring was based on the instances of clear and discrete concern spikes that have occurred over the data range (2010-2021). Over this 12-year period, three such instances occurred, that coincided with the severe UK flooding of 2014, Extinction Rebellion protests in 2019, and the start of COP26 in November 2021. From this, concern shocks were approximated as having a 3/12 or 25% chance of occurring on any given time-step in the model. The size of this effect was given by the average percentage change in concern between the start of a year and the point at which a spike occurred, which was calculated at 15%. If a concern shock occurs, the new agent influence weights feed through the model and agents make choices based on these updated values. At the end of the decision-making process and time-step (year), this effect is reversed to mimic the temporary nature of such concern shocks.

Scenario S6 follows a similar procedure, however, the probability is set a 1, to reflect the continuous nature of increasing concern. The size of this effect was modelled as the total annualized observed change in concern from 2010 to December 2021 (latest tracker data). To account for unequal distribution of tracker data, an effective daily value was calculated that was then annualized to give 1.65%.

Habit

In the model from Gibson et al. (2021), habit was treated as a multiplier to subsequent choice function scores that had repeatedly returned the highest value of the options available. That is, if a choice function of a given milk option consistently scored higher than the other option, eventually the habit bonus would trigger, further entrenching this option. We take the same form of this habit function, i.e. the empirical function of habit formulation is from Lally et al's. (2010) study of health behaviours remains the core component, however, in this study it was applied slightly differently. Here, it was additive rather than a multiplier, to ensure internal consistency with how the other four influence categories are modelled (physical, health, environmental and social). That is, a mixed additive and multiplicative weight and influence construct could yield disproportionate weight effects to their values. I.e. if one weight is added but another multiplied, this could increase or decrease their relative contribution, deviating from their assigned proportions. And so, to avoid this we followed a wholly additive approach. Further, its total impact is modulated by the weighting a given agent ascribes to the influence of habit on food choice. This is detailed by the following equation:

$$f(habit) = \beta_{hab} \cdot (peak\ habit - \exp(-0.042(consecutive\ choices - habit\ threshold))), \quad (4)$$

where 'peak habit' (fixed at two, but future model iterations should examine this with robustness analysis) is the maximum influence that habit can exert, 'consecutive choices' is the number of repeat highest scored milk choices across time-steps, and 'habit threshold' is the level that habit effects are triggered. The numerical value of the exponent (0.042) is directly from Lally et al. (2010) and β_{hab} is the weight assigned to habit.

Social influence

This sub-model represents the process of how agents influence, and are influenced by, other agents in their network (modelled as a small-world network (Watts & Strogatz 1998)), referred to here as an agent's neighbours. Note, this does not represent 'neighbours' in the strict geographical sense, but is inclusive of broad social interaction (e.g. family, friends, within households, local environment). The total number of nodes on the network is equal to the agent population. Each agent is initially connected to a number of neighbours set by the 'network-parameter', which can take an even integer value between 2 and 10. The rewiring probability is set at 0.1. Social networks also exhibit scale-free characteristics, and so a small-world scale-free network would perhaps give a more realistic representation. However, this network type is not available among the core set of NetLogo network extensions, which we acknowledge as a limitation of the study.

This study adapts the original formalism of social influence employed by Gibson et al. (2021). As with the original model, an agent has a probability of interacting, where influence is modelled as the mean set of choice functions across its neighbour network. However, instead of a free parameter that was termed 'social susceptibility', the extent of this neighbour influence is governed by the weighting an agent ascribes to social food influence (operationalised from BSA 2008 survey data). This is represented by the following equation:

$$f(social) = \beta_{soc} \cdot f(cog.)_{mean\ neighbour}, \quad (5)$$

where $f(cog.)_{mean\ neighbour}$ is the average value of neighbour cognitive choice functions and β_{soc} is the weight assigned to social influence.

Total choice function

The total choice function for each option is then given by the weighted sum of each influence component, expressed by the following equation:

$$
\begin{aligned}
f(total) &= f(cog.) + f(habit) + f(social) \\
&= \beta_{phy} \cdot (physical) + \beta_{hel} \cdot (health) + \beta_{env} \cdot (environmental)\,, \\
&\quad + \beta_{hab} \cdot (peak\ habit - \exp(-0.042(consecutive\ choices - habit\ threshold))) \\
&\quad + \beta_{soc} \cdot f(cog.)_{mean\ neighbour} \qquad (6)
\end{aligned}
$$

Evaluation

Agents have the opportunity to evaluate, learn from, and inform their future milk choices. This function remains largely intact from Gibson et al. (2021), employing a conceptualisation of cognitive dissonance between an agent's human values (from ESS 2018 survey data) and the impact of their milk choice behaviour (see Table 2). The minor update in this study is that agents now also look to minimise or escape a state of cognitive dissonance by altering the weight (+/-10% per time-step) they ascribe to health and environmental components versus physical (price) aspects. This is an effort to further draw on the empirical data from BSA 2008.

Table 2: Health and environmental impacts of dairy and plant-based milks. Values are weighted according to relative market/consumption shares of constituent products (e.g. whole, semi and skimmed for dairy, almond, soya and oat for plant-based). Sources for nutrition data; Vanga et al. (2018), Röös et al. (2018). Dairy GHG data are specific to the British Isles and from Clune et al. (2017). All other environmental data is sourced from Poore & Nemecek (2018).

Milk (per litre)	Sugar (g)	Sat. Fat (g)	Protein (g)	GHG ($kgCO2_{eq}$)	Land Use (m^2)	Water Use (L)
Dairy milk	50.32	9.78	36.81	1.12	9.00	628.00
Plant-based milk	31.99	1.96	13.75	0.85	0.64	174.73

Additional ODD elements

Emergence

The key results of modelled outputs that emerge from the behaviours and interactions of individuals are the macro-level average consumption of milk choice among the simulated population, the trajectories of these curves, and their proximity to delivering on dairy reduction target levels.

Interaction

Individuals interact with other individuals through a social network (small-world structure) where information exchange occurs. The mechanism by which information exchange occurs is not explicitly modelled, rather, this is governed by a probability of interaction and a social influence weighting

Stochasticity

Information on health and environmental impacts of different milk choice options that agents perceive is randomly drawn from a normal distribution with mean values. In the environmental concern simulation experiment, the occurrence of a public concern ‘shock’ is governed by a probability informed by observed data. Further, stochasticity is reflected in the logistic function of neighbour milk choice information entropy and probability of an agent becoming disposed to consider their alternatives.

Heterogeneity

Heterogeneity is represented by the assignment of state variables among the agents. Principally, agents have different milk choice influence weightings operationalised from British Social Attitude 2008 survey data, and different basic human values assigned according to the distribution of UK results data from the European Social Survey 2018.

Observation (incl. Emergence)

At each time step, the component choice functions and decision-making function for each choice and each agent is collected.

Appendix 1

Parameter	Description
1. Memory length	The size of an agent's memory that it can recall previous information. Cognitive perception is based on averaging values in the memory.
2. Habit threshold	The number of consecutive choices that return the same majority milk type consumption needed before the effects of habit take place.
3. Probability of interacting	The probability of an agent interacting (exchanging information on milk choice function scores) with other agents in its network.
4. Initial habit of incumbent	The initial number of consecutive choices that have returned the same majority milk type.
5. Social blindness	The probability that an agent has the ability to perceive the impact of its choice and therefore the option of evaluating it.
6. Post-choice justification	The threshold beyond which an agent will simply justify the discrepancy between its values and behaviour (milk choice impacts), rather than act to resolve it.
7. Cognitive dissonance threshold	The threshold below which any discrepancy between an agent's values and its behaviour (milk choice impacts) will not trigger a state of cognitive dissonance.
8. No. of neighbours	The number of neighbours in an agent's network.
9. Perception of health impact of PBM	The perception of the health impact of PBM.
10. Perception of environmental impact of PBM	The perception of the environmental impact of PBM.
11. Gradient of probability disposition	The slope of the function that determines how quickly the probability of being disposed to consider choice of milk as a function of the informational entropy of milk choices in an agent's neighbour network.
12. Perception of health impact of dairy	The perception of the health impact of dairy milk
13. Perception of environmental impact of dairy	The perception of the environmental impact of dairy milk.

Table A1: Model parameters and descriptions.

Influence	Category	Influence	Category
Quality/freshness	Physical	What family eat	Social
Taste	Physical	Recommendations	Social
Presentation etc.	Physical	Organically produced	Environmental
Availability	Physical	Animal welfare	Environmental
Price/value/special offers	Physical	Impact/fair trade/local	Environmental
Healthy/low fat	Health	Impact on landscape	Environmental
Vegetarian/special habits	Health	Packaging amount	Environmental
Additives/E-numbers	Health	Other	None
Habit/routine	Habit	Someone else decides	None
Try new/different	Habit	No particular	None
Know how to cook/prepare	Habit		
Convenient to prepare	Habit		

Table A2: Food influence response options in the BSA 2008 survey, and model categorisation.

Question	Survey	Model use
She/he strongly believes that people should care for nature. Looking after the environment is important to her/him.	European Social Survey, 2018	This question relates to the ‘universalism’ value and responses inform the environmental value position of agents used in evaluation sub-model.
It is important to him/her to live in secure surroundings. She/he avoids anything that might endanger his safety.	European Social Survey, 2018	This question relates to the ‘security’ value which contains the heath dimension. Note, the expanded 40 item PVQ includes a direct question on health, *‘She/he tries hard to avoid getting sick. Staying healthy is very important to her/him’*, but in the absence of this data in the ESS, we opt for the most relevant security value question.

Table A3: Questions from European Social Survey data used in the model.