The effectiveness of personalised food choice advice tailored to an individual's socio-demographic, cognitive characteristics, and sensory preferences

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.appet.2024.107600

Publisher: Elsevier

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The effectiveness of personalised food choice advice tailored to an individual's socio-demographic, cognitive characteristics, and sensory preferences

R.A. Briazu [a], L. Bell [a], G.F. Dodd [a], S. Blackburn [a], C. Massri [b], B. Chang [c], S. Fischaber [d], A. Kehlbacher [e], C.M. Williams [a], L. Methven [f], R. McCloy [a,*]

[a] *School of Psychology and Clinical Language Sciences, University of Reading, Berkshire, UK*
[b] *EU Collaborative Projects Area, European Food Information Council, Belgium*
[c] *Research Area, European Food Information Council, Belgium*
[d] *Analytics Engines, Belfast, Northen Ireland, UK*
[e] *German Aerospace Center DLR, Institute for Transport Research, Berlin, Germany*
[f] *Department of Food and Nutritional Science, University of Reading, Berkshire, UK*

ARTICLE INFO

ABSTRACT

Personalised dietary advice has become increasingly popular, currently however most approaches are based on an individual's genetic and phenotypic profile whilst largely ignoring other determinants such as socio economic and cognitive variables. This paper provides novel insights by testing the effectiveness of personalised healthy eating advice concurrently tailored to an individual's socio-demographic group, cognitive characteristics, and sensory preferences. We first used existing data to build a synthetic dataset based on information from 3654 households (Study 1a), and then developed a cluster model to identify individuals characterised by similar socio-demographic, cognitive, and sensory aspects (Study 1b). Finally, in Study 2 we used the characteristics of 8 clusters to build 8 separate personalised food choice advice and assess their ability to motivate the increased consumption of fruit and vegetables and decreased intakes of saturated fat and sugar. We presented 218 participants with either generic UK Government "EatWell" advice, advice that was tailored to their allocated cluster (matched personalised), or advice tailored to a different cluster (unmatched personalised). Results showed that, when compared to generic advice, participants that received matched personalised advice were significantly more likely to indicate they would change their diet. Participants were similarly motivated to increase vegetable consumption and decrease saturated fat intake when they received unmatched personalised advice, potentially highlighting the power of providing alternative food choices. Overall, this study demonstrated that the power of personalizing food choice advice, based on a combination of individual characteristics, can be more effective than current approaches in motivating dietary change. Our study also emphasizes the viability of addressing population health through automatically delivered web-based personalised advice.

1. Introduction

A diet low in low in fruit and vegetables and high in sugar and saturated fat is one of the leading risk factors for preventable ill health and premature mortality (Afshin et al., 2019). Adherence to generic healthy eating recommendations, such as the Eatwell guide, which offers suggestions regarding the amount and types of foods and drinks one should consume, has the potential to increase life expectancy and prevent cardiovascular diseases and some cancers (Cobiac, Scarborough, Kaur, & Rayner, 2016). However, generic dietary advice often has limited impact due to reduced awareness and low adherence. Goodman and colleagues report only a 18.2% recall of dietary guidelines from a sample of over 5000 participants, and more worryingly, in the UK less than 1% of people are achieving all of the Eatwell Guide recommendations (Steenson & Buttriss, 2021). To address this issue, personalised nutrition, an approach that uses information on individual characteristics to develop targeted nutritional advice, has been gaining ground. Personalised nutrition is based on the notion that tailored nutritional

advice, products, or services will be more successful at stimulating change than more traditional generic approaches (Jinnette et al., 2021; Ordovas, Ferguson, Tai, & Mathers, 2018). Indeed, Food4Me, the largest randomised controlled trial investigating the effect of personalised nutrition, found that tailored advice can help reduce red meat consumption on average by 8.5% and salt consumption by 6.3% when compared to population-based nutritional advice (Celis-Morales et al., 2017). Food4Me, as well as other studies investigating personalised nutrition, have specifically focused on tailoring advice to biological factors such as genotypic and phenotypic characteristics (Jinnette et al., 2021). However, food choice is a complex behaviour determined by a multitude of factors, therefore a more holistic approach to personalised nutrition is needed. For example, based on a review of conceptual models, Chen and Antonelli (2020), highlight that beyond biological factors an individual's food choice is also influenced by psychological factors such as personal identity, cognitive factors such as knowledge and skills, and socio-cultural factor such as economic variables. Additionally, environmental factors both social and physical well as food specific factors such as sensory and perceptual features are important. In the current paper we will seek to understand whether personalizing dietary advice according to a subset of these variables, namely cognitive biases, sensory preferences and socio-demographic characteristics can motivate healthier food choice.

Cognitive biases are systematic errors in thinking that result in deviations from rational decision making (Tversky & Kahneman, 1974). Several cognitive biases have been linked to dietary indices. For example, delay discounting, namely the tendency to prefer small immediate rewards over larger delayed payoffs (Frederick, Loewenstein, & O'donoghue, 2002) has been linked to purchasing a high proportion of foods from fast-food restaurants and increased BMI (Body Mass Index) (Appelhans, Tangney, French, Crane, & Wang, 2019). Similarly, a tendency to rely on quick intuitions (heuristics) rather than more consuming reasoning processes, a skill known as cognitive reflection (Frederick, 2005), is associated with higher caloric intake (Leitch, Morgan, & Yeomans, 2013). Currently, no personalised nutrition approach offers advice based on individual differences in cognitive bias susceptibility, however evidence suggests that such an approach should be considered. For example, interventions known to address delay discounting by highlighting the immediate benefits of a healthy diet, are more likely to encourage healthy eating (Satia, Barlow, Armstrong-Brown, & Watters, 2010). Similarly, those that are more likely to be biases by intuition are also likely to hold food heuristics (e.g. regarding healthy food as less filling), for these individuals highlighting the nourishing aspects of healthy food has been shown to lead to greater feeling of satiety (Suher, Raghunathan, & Hoyer, 2016).

An individual's sensory preferences, such as liking of specific food taste, texture and visual presentation also affects healthy food consumption, for instance a preference for sweet, salty, and fatty tasting foods can lead to the consumption of lower nutrient foods (Liem & Russell, 2019). Evidence suggests that taste-focus labelling of food rather than labelling based on health attributes are more likely to lead to an increase consumption of vegetables (Turnwald & Crum, 2019). Despite these findings, recommendations based on taste preferences have been largely overlooked in personalised dietary advice. There has been an effort to build recipe recommendation systems that align with a customer's flavour preferences (based on previous consumption), however there is limited application of these systems within a healthy eating context (e.g. Nag, Pandey, & Jain, 2017, October). Amiri, Li, and Hasan (2023) showed an automated meal-planning system that considered participants' taste preferences to be effective in addressing health-related nutrition intake, and evaluated by participants as helpful, however the impact of such recommendations on dietary behaviour change has yet to be evaluated.

Finally, socio-demographic characteristics such as age, income and household characteristics can influence dietary choice. Low-socio-economic status has been consistently associated with a diet low in

fruit and vegetables and high in refined sugar and saturated fat (d'Angelo, Guthrie, Draper, & Gloinson, 2020); for example, low -income households struggle to follow EatWell guidance (Scott, Sutherland, & Taylor, 2018) and tend to consume more take-away meals (Miller & Knudson, 2014). The link between low socioeconomic status and a diet low in fruits and vegetables and high red processed meats is unsurprising considering the relatively low cost of high sugar, high fat foods and relatively high cost of fruit, vegetables and animal protein foods in many countries (Headey & Alderman, 2019). Currently, most personalised nutrition interventions have the potential to exacerbate these health inequalities, given that such interventions are more accessible and more readily adopted by those with higher socioeconomic status (Pérez-Troncoso, Epstein, & Castañeda-García, 2021). To address this limitation, it has been recommended that personalised nutrition advice should be more easily available via the internet (Mathers, 2019). Furthermore, given that the cost is one of the biggest barriers to healthy eating in disadvantaged adults living in the UK (Briazu et al.,2024), personalised approaches should incorporate advice that is cost-effective for the consumer. Some personalised nutrition approaches such as the FoodSmart application (https://foodsmart.com/) have integrated budgetary considerations by allowing consumers to filter recommended recipes according to how budget friendly they are. The use of the FoodSmart approach has been shown to reduce food insecurity whilst also increasing diet quality (Bakre et al., 2022)

Equally, older age has been associated with nutrient deficiency (Clegg and Williams, 2018) and household structure, such as whether children are present in the household influences the ability to secure food and have access to a healthy diet (Caswell & Yaktine, 2013). Other than providing advice based on nutritional needs that differ by age, personalised approaches should also focus on adapting advice according to lifestyle. For example, the INCluSilver project specifically focused on providing personalised advice to older individuals highlights that systems developed to include advice regarding preparation approaches that can be physically managed by older adults, are anticipated to lead to better health outcomes (Burton, Wilmot, & Griffiths, 2018). Although it is well-documented how each of these factors can individually influence food intake, no study has yet investigated the effectiveness of addressing these factors simultaneously in a personalised approach. Therefore, personalised strategies focusing on a single dimension might conflict with other goals and constraints of the consumers (e.g., budget, time, sensory preferences, ability to understand the advice). Dietary recommendations that combine advice tailored to an individuals' cognitive, sensory, sociodemographic, and economic characteristics could be more motivating when compared to "one-size fits all" guidance. This hypothesis is in line with Social Cognitive Theory (SCT) that posits that emphasizes the dynamic interaction between personal factors, behaviour, and environments (Bandura, 1998). SCT-based interventions positively impact health outcomes and intervention effectiveness (Islam et al., 2023).

One innovative method that has recently been innovating decision-making within the field of healthcare is the use of synthetic datasets. Synthetic datasets represent "data that has been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s)" (Jordon et al., 2022). Use of synthetic datasets has a range of benefits including the ability to address data scarcity, privacy concerns, and decrease cost (Giuffrè & Shung, 2023). Within healthcare, synthetic datasets have been successfully applied to a diverse range of topics such as policy simulations (Davis, Lay-Yee, & Pearson, 2010) and diagnosis of COVID-19 based on CT scans (Das et al., 2022). Furthermore, synthetic datasets have been heralded as the most efficient way to enable new opportunities created by generative artificial intelligence, such as automated methods to provide personalised medical diagnosis and treatment (Chan, 2024; Omotunde & Mouhamed, 2023)

In the current paper we show how a synthetic dataset can be used within the context of personalised dietary advice and outline the process

used to develop and evaluate the effectiveness of such a personalised food advice strategy. The aim of targeting advice to several variables simultaneously is to enable individuals to align their food choice with their intentions by removing the cognitive load of processing these multiple factors. We use existing datasets to create a cluster model identifying groups of individuals for which we can personalise food choice advice, hypothesising that this will motivate change more than generic advice. We first created a synthetic dataset encompassing all variables of interest (Study 1a) and identified clusters of individuals with similar characteristics within this synthetic dataset (Study 1b). Finally, in Study 2 we evaluated the efficacy of food choice advice specifically targeted to some of these clusters, and assessed whether individuals are more likely to be motivated by these messages rather than the existing "one-size-fits all" government approach.

2. Study 1a – development of synthetic dataset

To the authors' knowledge, no interdisciplinary dataset of sufficient sample size combining information on cognitive, sensory, sociodemographic and economic variables currently exists. Here, we aimed to use existing datasets to create a new synthetic dataset, using a statistical matching technique, a procedure that estimates missing values while combining existing datasets that contain varying levels of overlap across variables of interest (European Commission, 2013; D'Orazio, 2019).

2.1. Method

2.1.1. Data selection

We initially identified accessible datasets that included information on UK participants for some or all the variables of interest, namely socio-demographic information, information on sensory preferences and cognitive biases, as well as information about food purchases. We sourced both publicly available datasets as well as datasets shared across departments at the University of Reading (UoR). In total 13 datasets that included UK adults (>18 years) were investigated, the majority from the Departments of Psychology, Food & Nutritional Sciences and School of Agriculture, Policy & Development at UoR. Additionally, we also investigated the publicly available dataset from the Living Cost and Food Survey in 2015 (LCFS), as this is one of the largest UK based datasets that provides essential information for key social and economic variables. Each data set was evaluated according to the following criteria: sample size, representativeness of the UK population and compatibility between variables in terms of the constructs measured. Following this process three datasets were selected as sources for the synthetic data set.

The LCFS provided survey data from over 5000 households representative of the UK population including (but not limited to) information on food purchases collected via shopping receipts every two weeks and sociodemographic characteristics (Office for National Statistics, 2019). The second dataset was collected as part of the project entitled 'Cognitive Biases and Behavioural Segmentation in Food Demand' completed by the University of Reading, hereinafter referred to as CogSeg. This dataset contains data from 732 participants featuring cognitive, economic and socio-demographic factors, alongside food purchase information. Purchased food items were matched to former UK Ministry of Agriculture, Fisheries and Food (MAFF) codes to help categorize food items and high-level MAFF codes summarised purchases into 1) milk & dairy, 2) meat & fish, 3) cereal, fat, etc, 4) fruit & veg and 5) drinks (Smithers, 1993). The third dataset (referred to as SenseSeg) contains socio-demographic, economic, food purchasing, cognitive as well as sensory preferences for foods characterised by different tastes, and a

Table 1
Variables measured within each dataset. Dark grey cells indicate that data for that variable was present for all participants, light grey indicates data was present only for some participants, white cells indicate data for those specific variables was not included in the dataset.

Variables		LCFS (n=3654[1])	CogSeg (n=732)	SenseSeg (n=600)
Demographic	Gender			
	Age			
	Household size			75% missing
	BMI			
Economic	Household income			
	Household reference person income			
	Employment status			
Food purchases	Food expenditure per person per week			75% missing
	Proportion of fruit and veg purchased			75% missing
Cognitive	Mental accounting			
	Cognitive reflection			
	Discounting			
	Sunk Cost			
Sensory	Food Neophobia			
	Sensory liking for salty taste			
	Sensory liking for bitter taste			
	Sensory liking for sweet taste			
	Sensory liking for fatty sweet taste			
	Sensory liking for fatty salty taste			

[1] This is the total number of households for which there was no missing values.

measure for food neophobia from 600 UK participants. The sensory survey was based on a UK version of the French PrefQuest questionnaire (Deglaire et al., 2015) validated at the University of Reading (see Appendix A for details about how these scores were calculated and follow the on-line link for details about survey items for each taste preference: https://osf.io/47xem. Table 1 shows the variables measured in each dataset.

2.1.2. Procedure to create the synthetic dataset

We used the micro approach of Statistical Matching (European Commission, 2013) to produce a synthetic data set with entries for all households in the LCFS data set that had data for all variables of interest (namely, demographic, economic and food purchase data), as this was the largest and most UK representative dataset for the UK.

To ensure that the format for all variables matched, we transformed all variables into categorical variables. Categorial format for each variable is presented in Appendix A. We used the Hellinger Distance metric (European Commission, 2013) to determine the similarity between the distributions of variables in all datasets, with two distributions classed as being similar if the Hellinger Distance is 0.05 or smaller (European Commission, 2013). Typically, the variables in our data sets had larger Hellinger Distances. For example, the Hellinger Distance for gender is 0.16 between LCFS and ESRC and 0.24 between LCFS and Sensory. To account for this, we assessed several different statistical matching methods and found that the Dirichlet Process Mixture of Products of Multinomials (DPMPM) (Hu, Reiter, & Wang, 2018) was the most appropriate matching process for our data, due to its ability to cope well with lower degrees of closeness of common variables.

The DPMPM method treats statistical matching as a missing data problem. The model finds groups of individuals with similar characteristics through an iterative Markov Chain Monte Carlo (MCMC) procedure. Each group of similar individuals in the data can be characterised by a set of parameters capturing the joint distribution of category values across all variables. This can then be used to sample from and obtain values to impute missing observations. Accuracy of the synthetic data set is improved by using auxiliary information, called glue (Fosdick, DeYoreo, & Reiter, 2016). Glue is a way to add additional information to the statistical matching and can act like a prior distribution. Studies show that the use of glue improves the results of statistical matching as it can inform the overall distribution of a variable when the data are skewed (Fosdick et al., 2016). For example, in our data, gender is distributed differently across the three data sets. Adding the 'true' distribution from the LCFS as glue fixes the posterior distribution to this distribution while not impacting any of the other relationships between variables and gender.

2.2. Results and discussion

The resulting synthetic dataset included data from 3654 households with information on socio-demographics and economic characteristics from the LCFS data and statistically matched data on sensory and cognitive variables. Fig. 1 shows the input datasets and the makeup of the resulting synthetic dataset.

Demographic information included: age, gender, household size and BMI (calculated based on self-reported height and weight). Economic variables referred to income as indexed by yearly gross household income and employment status (whether individuals were in full-time employment or not). Weekly food expenditure per person was calculated by dividing total expenditure by number of household members. This alongside percentage of purchases of fruit and vegetables represented food purchase information. In terms of sensory variables, the dataset included information on participants' food neophobia (i.e. an individual's tendency to avoid unfamiliar food) and liking of foods characterized by salty, bitter, sweet, fatty sweet and fatty salty tastes (see Appendix A for details about how these scores were calculated and follow the on-line link for details about survey items for each taste

reference; https://osf.io/47xem). Data on food neophobia was collected using the validated 10 item questionnaire (Pliner & Hobden, 1992). The sensory taste liking questionnaire was developed at UoR and adapted from the French PrefQuest questionnaire (Deglaire et al., 2015). The 8 to 19 food items per taste (sweet, bitter, salty, fatty-sweet and fatty-salty) were based upon commonly consumed UK foods, and the questionnaire was validated by both test-retest and against liking ratings of real foods representing the listed food in the sensory laboratory at UoR (data not shown).

Lastly, the dataset included information on the extent to which participants showed each of four cognitive biases, namely (i) delay discounting defined as an individual's strong preferences for small immediate payoffs relative to larger delayed payoffs (Frederick et al., 2002); (ii) cognitive reflection, the capacity to override an incorrect 'gut' response in favour of a correct response that requires deliberation (Frederick, 2005); (iii) mental accounting described the process whereby people code, categorize and evaluate economic outcomes (Kahneman & Tversky, 1984), and (iv) resistance to sunk costs which refers to the ability to ignore prior investment when making decisions (Arkes & Blumer, 1985). Information on the way each variable was categorized within the synthetic dataset can be found in Appendix A.

All of these variables have been found to influence food choice. Low-socio-economic status has been consistently identified as a risk factor for a diet including more foods with a low nutrient density diet (d'Angelo, Guthrie, Draper, & Gloinson, 2020) and older age has been associated with nutrient deficiency (Clegg and Williams, 2018). Socio-economic status affects diet through food purchasing because one of the largest barriers to achieving healthier eating habits for people with lower income is the cost of healthy food (Jones, Tong, & Monsivais, 2018; Scott et al., 2018). Taste is another main factor that influences food choice. Research suggests that individuals tend to focus on taste rather than health (Roininen, Lähteenmäki, & Tuorila, 1999), and this often leads to increased consumption of foods with lower nutrients (Liem & Russell, 2019). A study including over forty-six thousand French adults found that liking for salt and fat was positively associated with BMI (not sex specific), as was liking for sweet foods with women (Deglaire et al., 2015). In terms of cognitive biases, a tendency for delay discounting has been linked with higher caloric intake (Appelhans et al., 2012) and fast-food purchases (Appelhans, Tangney, French, Crane, & Wang, 2019). Those subject to the sunk cost fallacy are likely to continue eating after feeling satiated (Jarmolowicz, Bickel, Sofis, Hatz, & Mueller, 2016) whilst a tendency for reflection impulsivity (as indexed by low cognitive reflection scores) is significantly associated with uncontrolled eating and thus weight gain and a BMI over 30 (Leitch et al., 2013). Finally, mental accounting may dictate the way in which individuals choose to purchase food items (Milkman & Beshears, 2009). Importantly, identifying how the different categories of these variables cluster together to define different types of individuals is necessary to provide personalised food choice advice. Therefore, we next aimed to identify such clusters by using probabilistic clustering methods.

3. Study 1b – creating the cluster model

Having created a dataset that includes all the variables of interest, the next step was to identify groups of individuals within this dataset with similar characteristics. In this study we describe the methods used to identify how individuals from our synthetic dataset cluster together based on all 18 individual variables as identified in Study 1a. We then describe how socio-demographic, economic, cognitive and sensory characteristics vary across the clusters. Finally, we provide a more in-depth description of a selection of clusters that we intend to target when designing personalised food choice advice.

3.1. Method

Similarly to the statistical matching process for the synthetic dataset,

			Variables			
Data sets	observations	Demographic	Economic	Food purchase	Cognitive	Sensory
LCF	3654				imputed	imputed
ESRC	687					
Sensory	512			25% missing		
Glue ESRC	687					
Sensory	512					

Fig. 1. Statistical matching data input and output. Grey tiles indicate observed variables for each given data set.

a Dirichlet Process Mixture Model of Multinomial Products (DPMPM) model was estimated using the R-package PReMiuM (Liverani, Hastie, Azizi, Papathomas, & Richardson, 2015).

Dirichlet Process mixture models provide a way of clustering data without a pre-defined number of clusters. The assumption is that we have observed a finite but unbounded number of clusters in our data from a wider population that is believed to have an infinite number of unobserved classes (Heinz, 2015). Thus, there is always a positive probability that the model assigns a new observation to a new cluster. Following (Heinz, 2015), a Dirichlet Process mixture model can be summarised by Equation (1) where π denotes the mixing proportions, θ_k represents the class parameters and z_i the latent class assignment for household i to class k. The model assumes that each of n observations of $x_i \in \{x_1, ..., x_n\}$ is assigned to one of K classes where $K \in \{k_1, ..., k_\infty\}$. Individuals get assigned to class z_i with probability of class proportions π_k, where $Mult(\pi)$ is the multinomial distribution satisfying $\mathbb{P}(z_i = k) = \pi_k$, and identified by cluster specific distribution $F_k = F(\bullet|\theta_k)$. Class parameters θ_k are drawn from a prior distribution $H(\lambda)$ and mixture proportions π are determined by precision parameter α through the stick-breaking process (Sethuraman, 1994).

An advantage of the model is that it estimates jointly the cluster parameters as well as the number of clusters, thereby accounting for the uncertainty in both. The model was estimated with 20,000 MCMC iterations and a burn-in of 10,000. The prior parameter α influences the scale of the Dirichlet process and therefore the number of clusters with a larger value resulting in more clusters. Setting $\alpha = 0.1$ resulted in the posterior mean for the cluster number $k = 27$ with a 90% credible interval of [27,28]. The optimum cluster allocation was identified using the Variation of Information (VI) approach (Wade & Ghahramani, 2018) as part of the R-package mcclust.ext (Wade S. , 2015) which compares any two clusters in terms of shared information and information within each cluster. Using the 20,000 posterior draws for cluster membership z_i, it computes the pairwise probability of all observations being in the same cluster. The optimal cluster assignment z_i^* is representative of the posterior and for each observation minimises the posterior expected loss of choosing z_i^* compared to the other 19,999 drawn z_i (*Wade & Ghahramani*, 2018). Using this approach, we identified the optimum draw $\theta_{k^*}^*$ where $k^* \in \{1, ..., 27\}$.

3.2. Results and discussion

The model found a total of 27 individual clusters (labelled 0 to 26) within the synthetic dataset which could be described in terms of differences in cognitive, sensory, food purchasing, demographic and

economic characteristics. Each cluster's characteristics are presented in. Appendix B Fig. 2 shows the number of individuals assigned to each cluster. The cluster with most individuals was cluster 6, which included 341 cluster members, the least members were assigned to cluster 26 that only included just one member, followed by cluster 18 with 39 members.

Income was the variable that most clearly drove cluster membership, meaning that within each cluster most participants belonged to a single income category. In seven clusters, most members (over 88%) reported a low income, whilst in five clusters most members (over 95%) reported a high income.

All clusters[1] included both males and females, and only in two clusters the vast majority of participants were female (over 68%; clusters 22 and 25). Only one age category, namely 65+, was a strong characteristic of five clusters (clusters 1, 6, 8, 11, and 21) whilst the remaining clusters included individuals from all age groups. In terms of household size, two clusters were entirely made up of individuals living on their own (clusters 16 and 18), whilst the remaining clusters were a mixture of different household sizes, although two member households were
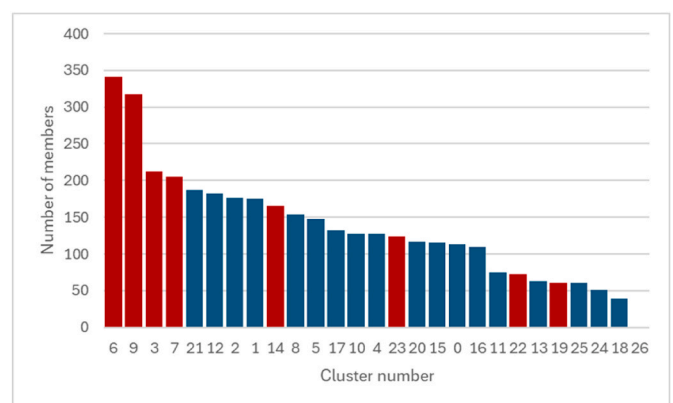

Fig. 2. Number of individuals in each cluster resulting from the cluster model arranged in descending order of cluster size. Cluster highlighted in red were targeted for further evaluation in Study 2.

---

[1] In this description of clusters, we do not include cluster 26, this cluster contained 1 individual, thus there was no variability in the levels of any variables.

predominant in three of the clusters (clusters 0, 11 and, 21). Based on BMI, two clusters included mostly members with a BMI over 30 (clusters 11 and 16), whereas another two included individuals with a low BMI, indicative of being underweight (clusters 9 and 19).

For food purchases, cluster 0 alone had most members spending a large amount on food, the income was also high in this cluster, although the household size was small. With regard to percentage of fruit and vegetable purchases only, cluster 18 had a particularly high spend on fruit and vegetables and only cluster 24 had a low spend (i.e. under 10%).

In terms of cognitive biases, one bias that was strongly present in several clusters was delay discounting. High delay discounting characterised over 80% of participants in four clusters (clusters 3, 10, 22, and 23). Low cognitive reflection, as indicated by an intuitive response on all items of the Cognitive Reflection Test (CRT), characterized over 70% of individuals in three clusters (clusters 1, 13, and 18). Sunk cost and mental accounting biases were less prevalent. A high sunk cost bias characterized just over 65% of individuals in one cluster (cluster 13). Similarly, the highest proportion of individuals with a high mental

accounting bias was 56% and this was in a cluster with a small number of members (cluster 18).

In relation to sensory linking, most clusters were not strongly characterized by only one taste preference, conversely, three clusters were characterized for strong preferences for all five types of taste (clusters 2, 13 and 15). Preference for a sweet taste was always accompanied by a preference for either fatty-sweet or fatty-salty taste (e.g. clusters 0 and 5). Three clusters showed a strong preference for the bitter taste (clusters 21, 17 and 19) and three clusters showed a strong preference for fatty related tastes (clusters 22 and 23 and 18). Very high neophobia characterized one cluster (cluster 1).

The characteristics of each cluster provide a basis for the development of cluster-specific messages on healthy eating. Our next aim was to understand whether framing dietary messages according to cluster characteristics could motivate positive dietary change. A subset of 8 clusters was selected with this aim in mind, as highlighted in Fig. 2. First, we chose clusters 6, 9, 3 and 7, as these were the four largest clusters and contained data from over 1000 individuals constituting about 30% of the synthetic dataset sample, meaning these clusters were the most
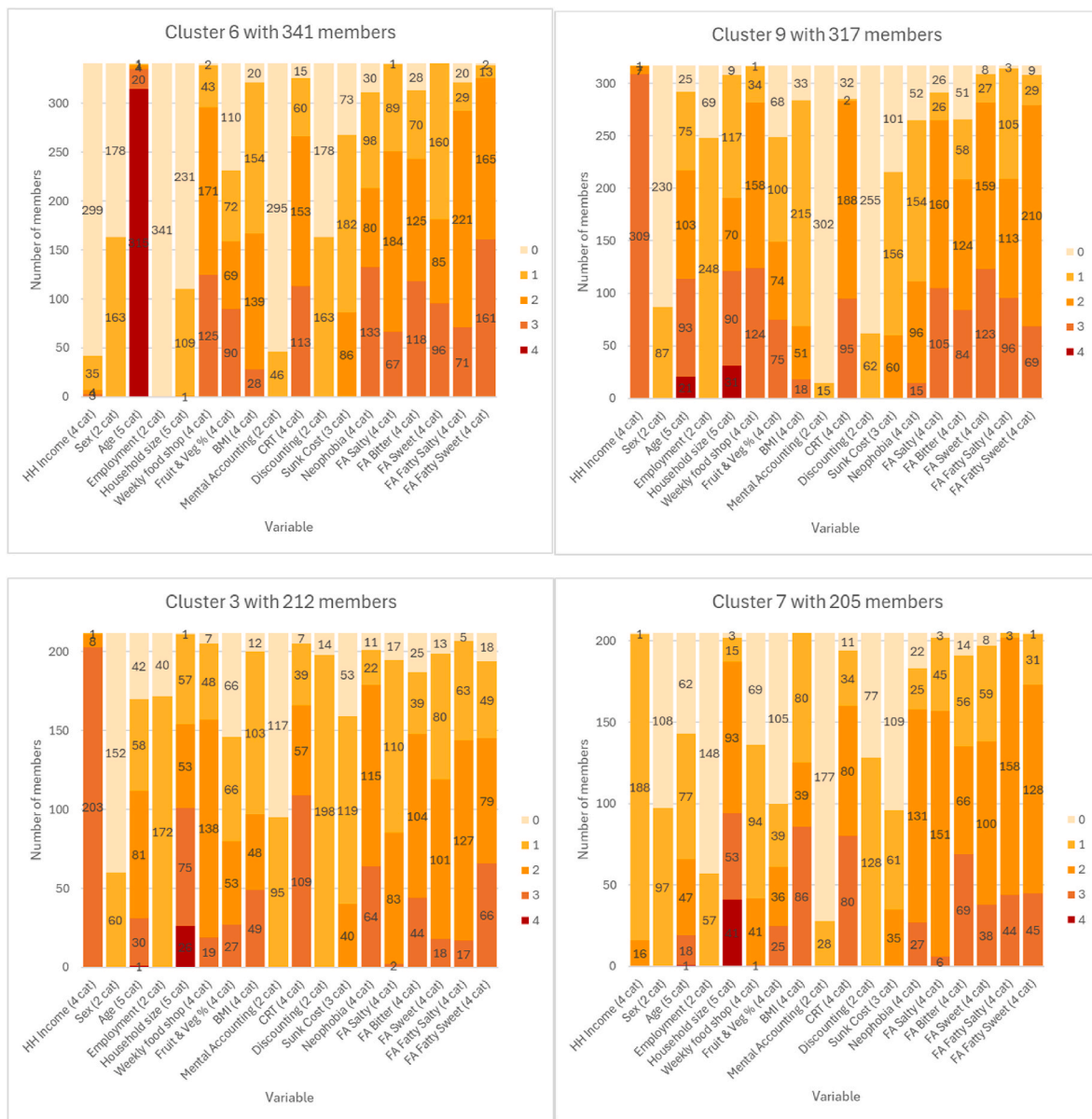

Fig. 3. Characteristics of the top four largest clusters. Each bar indicates how individuals are distributed along the categories of each variable. Category values for each variable can be found in Appendix A.

common in the population. Cluster 6 mostly included older adults, not in full time employment, on a low income who had a slight preference for fatty-sweet foods. In Cluster 9 most individuals had a higher income, lived with their families, displayed a reflection impulsivity and a resistance to discounting bias, with a preference for sweet foods. Cluster 3 was characterized by those in full-time employment, with a high household income. Individuals in this clusters are also likely to prefer small immediate rewards over bigger delayed pay offs (high discounters) and were reluctant to eat novel or unfamiliar foods (neophobia). The fourth largest cluster, cluster 7, was characterized by low income, larger households, high CRT bias and moderate neophobia. Characteristics of these top four clusters are shown in Fig. 3.

Second, given that the main novelty of our approach was to address cognitive and sensory variables within personalised advice, we chose a further four clusters that were each strongly characterized by either a sensory or cognitive variable, these are shown in Fig. 4. This enabled us to directly test our hypothesis. Taste preferences had not been a clear characteristic of the largest clusters; where there was a preference for certain tastes this was not particularly strong. For this reason, we chose to select two clusters with very strong preferences for one taste. Cluster

22 (n = 72) was chosen as this was the only cluster in which all members of the cluster showed a very strong preference for one taste, namely fatty-sweet taste. This cluster was also characterised by low income and a discounting bias. We also chose cluster 19 (n = 61) because individuals in this cluster included a high percentage of individuals with a preference for bitter. The reason for choosing this cluster and not the other two clusters with a strong preference for bitter, was because in this cluster there was also a very low percentage of bitter dislikers. Most individuals in cluster 19 were also earning lower incomes.

Additionally, we selected cluster 23 (n = 124) as this cluster showed the strongest discounting bias, it was also characterized by low income, and a preference for fatty-salty taste. Finally, we also included cluster 14 (n = 166), which had a high percentage of individuals with reflection impulsivity, in addition to multiple member households and a preference for fatty-sweet taste.

To reduce complexity, at this stage, we chose not to focus on sunk cost, mental accounting, food purchase information or BMI, as the discriminatory potential of these variables was low. The next step was to formulate advice based on the characteristics of each cluster and seek to understand the effectiveness of such advice in creating a change in diet.
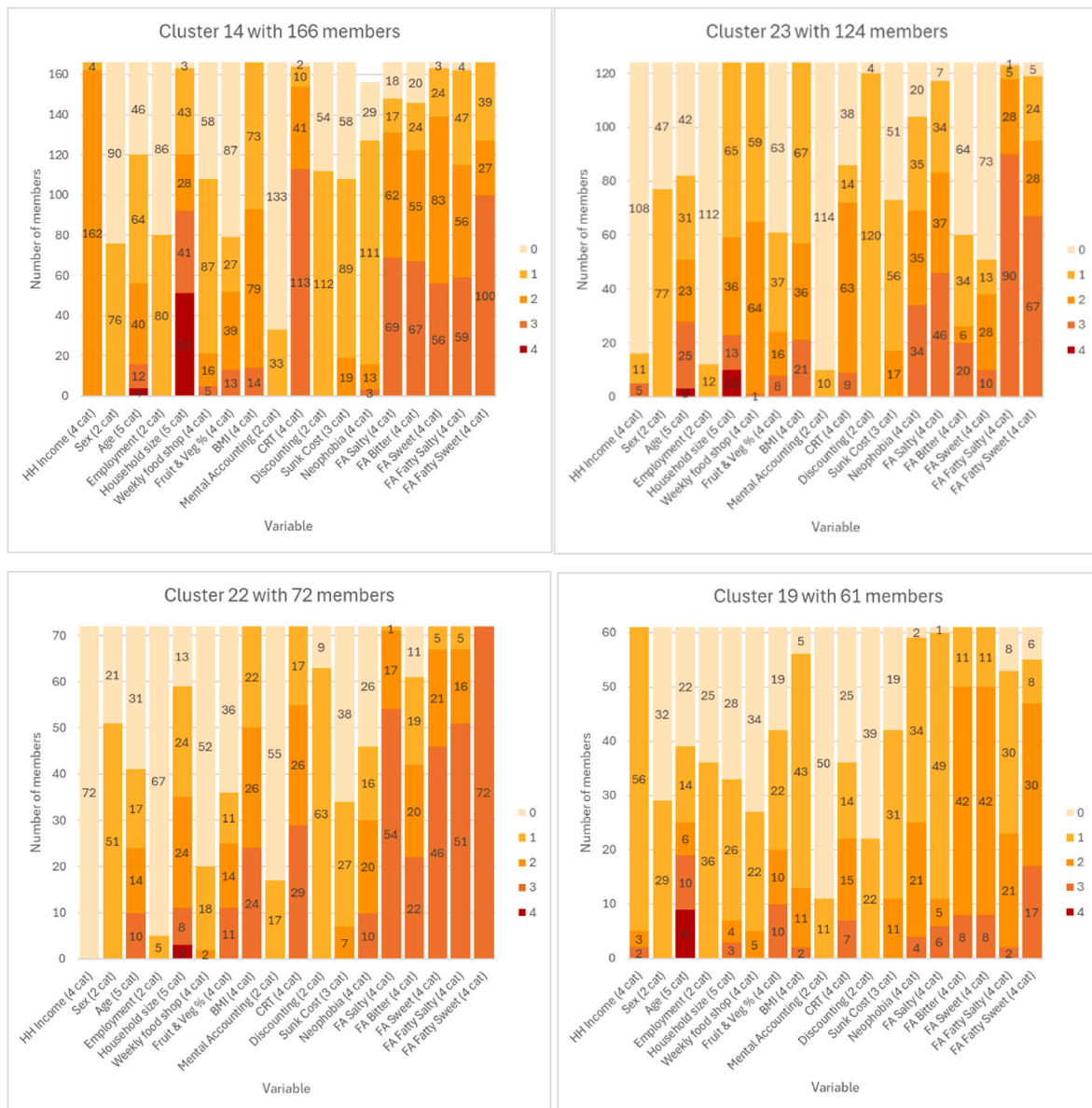

Fig. 4. Additional clusters chosen for further investigation in Study 2.

4. Study 2 – evaluation of personalised food choice advice based on cluster characteristics

The creation of a synthetic dataset, encompassing different variables associated with food choice, informed the subsequent cluster model and enabled the identification of unique profiles of groups of individuals, characterized by variation in food choice attributes. Understanding how the characteristics of each cluster are associated to food consumption provides an insight into how dietary messages could be framed to motivate healthy eating. For example, delay discounting bias can be overcome by clearly highlighting immediate rewards (Kurth-Nelson, Bickel, & Redish, 2012). In the context of food intake this could imply highlighting the immediate reward of consuming fresh fruit and vegetables (e.g., feeling energized and refreshed) rather than the delayed rewards (e.g., better health). Equally, socio-demographic characteristics such as employment status could be targeted by considering the barriers to healthy eating that individuals in such circumstance might experience. For example, we know that individuals in full-time employment can suffer time scarcity, and thus experience an increase in the consumption of fast food and ready-prepared meals (Jabs & Devine, 2006). Thus, emphasizing how dietary guidelines could be achieved through preparation short-cuts could be effective. Similarly, highlighting cheaper alternatives for those in low-income households could also be effective. Additionally, taste preferences can be addressed by recommending healthy alternatives that satisfy such preferences.

In this study we seek to build food choice advice specifically targeted to characteristics of a subgroup of clusters and assess the impact of such advice on motivation to change. We hypothesize that participants will be more motivated to follow dietary recommendations after reading the personalised advice rather than generic advice, but only if the advice matches the cluster they belong to.

4.1. Method

4.1.1. Participants

In part 1 a total of 390 participants took part in the study. Only 250 (64.10%) participants returned in part 2, however 33 (13.2%) participants were removed due to poor answers on an attentiveness check question. Of the remaining 218 participants, 59 participants were part of the control condition, 97 participants part of the unmatched personalised condition and lastly 62 participants part of the matched personalised condition. A post-hoc power analysis conducted using G*Power version 3.1.9.7. (Faul, Erdfelder, Buchner, & Lang, 2009), based on based on a medium effect size of 0.25 revealed that statistical power was 0.91 and therefore more than adequate to test the study hypothesis.

Demographic information (see Table 2) was collected using categorical questions to match the way in which variables were classified in the cluster model.

4.1.2. Procedure

This was a two-part online study. In the first part participants completed a questionnaire assessing demographic characteristics, cognitive variables, food intake and dietary preferences, including sensory taste preferences. Responses were used to identify the cluster participants belonged to and therefore informed the personalised food advice based on the characteristics of the cluster.

Two weeks after completing the questionnaire participants were recontacted and asked to read a paragraph containing food choice advice. Participants in the control condition read current UK food-based dietary guidelines about the consumption of fruit, vegetables, sugar and saturated fat. A second group received personalised advice according to the cluster they belonged to (matched personalised condition). Finally, the third group received personalised advice that was not matched to their cluster (unmatched personalised). After reading the advice, all participants had to rate how personalised they deemed the message to be, how adequate the recommendation to modify their diet was and the

Table 2
Demographic characteristics of sample (N = 281).

Characteristic	Total	%
Gender		
Male	133	61.01
Female	85	38.99
Age		
18-24	44	20.18
25-34	21	9.63
35-44	30	13.76
45-54	51	23.39
55-64	36	16.51
65+	36	16.51
Income		
Under £20000	61	27.98
£20000 – £29999	55	25.23
£30000 - £49999	43	19.72
Over £50000	51	23.94
Unsure/Missing	8	3.67
Household size		
1	47	21.55
2	95	43.58
3	28	12.84
4	32	14.68
5+	16	7.34

degree to which they anticipated they would change their consumption.

4.1.3. Materials

A shortened version of the original questionnaire items from the synthetic dataset was used. The rationale for shortening the questionnaire was two-fold. First, we wanted to account for respondent fatigue, second the aim was to create an application for digital devices thus we wanted something simple and easy to use. In a separate study, each set of items was subjected to psychometric analysis to ascertain reliability (information on this analysis can be found online https://osf.io/ns6tx). The way in which each variable was assessed can be found in Appendix A. The full questionnaire that participants filled in can be found online at https://osf.io/d4t8k.

Dietary intake was also measured in order to control for participants' current diet. Using a self-report measure we assessed fruit and vegetable consumption, saturated fat and sugar intake as follows: For fruit and vegetables we asked participants to report the number of fruit and vegetable portions consumed per day. Example of different sources of fruit and vegetables as well as quantities to signify what we meant by 'a portion' were provided. To ascertain saturated fat intake participants were asked to state how many times per week they cook with either lard or butter, how many times per week they consumed fatty meats (either cooked by themselves or in ready meals) and how many portions of high saturated fat dairy products such as cheese they consumed per week. Intake of sugar was ascertained by asking participants to state how many teaspoons of sugar they added to their drink in a day and how many fizzy drinks they consume per week. We also asked how many high fat and high sugar snacks (such as chocolate) participants consumed per week. This question targeted both saturated fat and sugar intake. For each of the questions asking frequency of consumption per day participants had to pick one of five options, namely None/Never, 1–2, 3–4, 5–6 or 7+. Questions asking participants to state consumption per week were provided with the following options: None/Never 1–4, 5–8, 9–12 13+.

4.1.3.1. Advice messages.
Each advice message was divided into several sections each dedicated to advice concerning consumption of fruit, consumption of vegetables, advice regarding sugar and finally advice about the consumption of saturated fat.

The control message was based on the NHS EatWell Guide (Buttriss, 2016). This message included information regarding the minimum number of fruit and vegetables portions to be consumed per day, along-side the weight of what was meant by 'a portion' and different

sourcing options. Guidelines regarding the intake of dried fruits was also provided. For sugar consumption, the paragraph read by the control group advised on the maximum intake per day and the option to replace sugar with fruit. Lastly for saturated fat, participants were provided with the maximum quantity they should consume per day depending on age and gender (see Appendix C).

For the personalised messages, the control paragraph was used as a template that could be changed or modified in accordance with the characteristic of each cluster. All messages included the same advice, namely to increase the consumption of fruit and vegetables, and decrease the consumption of sugar. Additionally, the message also stated the recommended daily portion for saturated fat depending on gender and age. However, the way in which these recommendations could be achieved was personalised based on each cluster's characteristics.

Low cognitive reflection was addressed by busting food myths and by providing an easy "rule-of-thumb" to aid healthy eating, thus creating healthy heuristics for individuals to rely on (Schulte-Mecklenbeck, Sohn, de Bellis, Martin, & Hertwig, 2013). Specifically, advice for clusters with individuals engaging in heuristics emphasized that frozen vegetables are just as nutritious as fresh ones, and that an easy way to include more fruits and vegetables in the diet would be to buy mixed bags of frozen fruit and vegetables and prepare soups or smoothies.

Preference for immediate rewards rather than larger delayed rewards was addressed by emphasizing the immediate benefits of consuming fresh fruit and vegetables. The advice stated that consuming fresh fruit can make one feel immediately refreshed and energized. Conversely, for cluster 9, characterised by individuals preferring delayed rewards the advice stated that eating a variety of fruit and vegetables could lower the risk of future illness.

Advice for clusters characterized by low income emphasized that tinned or frozen alternatives could be cheaper. For clusters with older participants, those in full time employment or belonging to larger families, advice emphasized how preparing healthy foods can be convenient. Ease of preparation was highlighted by mentioning the potential to increase intake of fruit and vegetables via soups or smoothies. Preparation short-cuts such as using tinned or frozen ingredients were also included.

Sensory preferences were addressed by providing fruit and vegetable suggestions. For saturated fat we also exemplified how potentially preferred options (e.g. fried chips for fatty-salty likers) could be replaced with healthier options with a high probability to appeal to the sensory liking of the cluster (e.g. baked potato with roasted garlic). Similar examples were provided for sugar alternatives. Neophobia was addressed by providing examples of the most commonly purchased fruit and vegetables such as bananas and apples and carrots and courgettes, based on the analysis of LCF data (Office for National Statistics, 2017).

As an example of the personalisation process, cluster 6 included older adults (in our cluster model over 93% of participants in this cluster were over 65 years of age) with a low income (nearly 88% earned below £20,000) with a preference for fatty-sweet taste. As such the advice we generated mentioned cheaper healthy alternatives such as tinned vegetables and emphasized how these are also easy to prepare and cook. Food alternatives mentioned in the advice were ones that we had previously been found to correlate the highest with a preference for fatty-sweet taste such as mango and melon, butternut squash and beetroot. Appendix D includes details of how personalised messages were tailored to each cluster. Full personalised advice for each of the 8 clusters can be found at: https://osf.io/47xem.

4.1.3.2. Materials for the evaluation of nutritional messages. After reading the advice messages, participants in all conditions were asked to state how personalised they felt the advice was, using a 6-point Likert scale that ranged from 'Not at all personalised' to 'Very personalised'. Participants in the matched and unmatched conditions were also asked to rate how appropriate they felt the recommendation to increase fruit

and vegetable consumption was, and how appropriate the recommendation to decrease consumption of sugar and saturated fat was. Participants in these two conditions were also asked to evaluate their anticipated sensory liking of the alternative food options offered using a 9-point Likert scale ranging from 'Dislike extremely' to 'Like extremely'.

Participants were also asked to rate how likely they were to change their consumption of fruit, vegetables, sugar and saturated fat based on the advice they had read (e.g. How likely are you to change your consumption of saturated fat based on your advice above?). For each question participants used a 6-point Likert scale ranging from 'Extremely unlikely' to 'Extremely likely'.

4.2. Results

4.2.1. Evaluation of messages

Table 3 shows that participants in the personalised condition rated their sensory liking for the sugar alternative in the advice significantly higher than participants in the non-personalised condition. However, participants in the two conditions did not differ in their sensory liking for the fruit, vegetable and fat alternative presented in the advice.

To understand whether participants across the three conditions differed in their perception of how personalised their message was (scored out of a maximum of 6), we performed a one-way ANOVA. As anticipated, participants in the control condition rated the message to be the least personalised (M = 2.81, SD = 1.36), followed by participants in the unmatched personalised condition (M = 3.84, SD = 1.44), and lastly those in the matched personalised condition (M = 3.92, SD = 1.37). This difference in how personalised the participants felt the message was significant between at least two groups (F(2,214) = 12.34, p < 0.001). Bonferroni post hoc tests found that participants in the control condition felt the message was significantly less personalised as compared to both the unmatched personalised group (p < 0.001, 95% [−1.47, −0.40]) and the matched personalised group (p < 0.001, [−1.73, −0.56]). However, the difference in ratings between the two types of personalised message conditions was not significant (p = 0.941, 95% [0.40, 1.47]).

Across both unmatched and matched conditions, there was a significant positive correlation between the anticipated sensory liking (all scored on a scale from 1 to 9) of the recommended fruit ($r_{(187)}$=0.34, p < 0.001), vegetables ($r_{(187)}$=0.41, p < 0.001), sugar alternative ($r_{(187)}$=0.47, p < 0.001), fat alternative ($r_{(187)}$=0.47, p < 0.001 and participants perception of how personalised the advice was.

For both unmatched and matched conditions all participants were specifically advised to increase their consumption of fruit and vegetables and decrease their consumption of sugar and saturated fat. We therefore wanted to understand whether participants' interpretation of how adequate this advice was related to their consumption. There was no

Table 3
Means and standard deviations (SD) and *t*-test results for sensory liking of food options offered as alternatives in the advice messages for each personalised condition.

	Unmatched personalised		Matched personalised		*t*-test	Significance
	Mean (1–9)	SD	Mean (1–9)	SD		
Anticipated sensory liking of fruit	6.30	1.99	6.47	1.78	−0.54	0.587
Anticipated sensory liking of vegetable	6.17	1.99	6.48	1.52	−1.07	0.287
Anticipated sensory liking of sugar alternative	5.30	2.18	6.15	1.70	−2.59	0.011[a]
Sensory liking of fat alternative	5.61	2.0	6.10	1.73	−1.58	0.116

[a] p < .05.

relationship between participants consumption of fruit ($r_{(190)}$ = 0.6, p = 0.417), vegetables ($r_{(190)}$ = −0.01, p = 0.877) and saturated fat ($r_{(190)}$ = 0.08, p = 0.284) and their perception of adequacy regarding the advice to increase or decrease their consumption of these diet components. However, the more sugar participants reported consuming, the more adequate they thought the advice to decrease their sugar consumption was ($r_{(190)}$ = 0.19, p = 0.008).

4.2.2. Intentions for diet change

Means and standard error for participants' willingness to change their intake of fruit vegetables, sugar and saturated fat (all scored on a scale from 1 to 6) across the three conditions are shown in Fig. 5.

One-way ANCOVAs were conducted to compare intention to change across all conditions whilst controlling for current intake. There was a significant difference in intention to change fruit consumption between conditions ($F(2,217)$ = 7.36, p = 0.001). Post hoc tests revealed there was a significant difference between the control condition and the matched personalised condition (p = 0.001 [−1.58, −0.63]). However, intention to change the consumption of fruit did not differ between the control and the nonmatched personalised conditions (p = 0.60 [−1.09, 0.02] or between the unmatched and matched personalised condition (p = 0.168, [-0.11, 0.97]).

There was also a significant difference in intention to change vegetable consumption between conditions whilst controlling for current vegetable intake ($F(2,217)$ = 12.26 p < 0.001). Bonferroni adjusted post-hoc tests revealed there was a significant difference between the control condition and both unmatched personalised (p = 0.004, [−1.30,-0.18]) and matched personalised (p < 0.001, [1.89, 0.64]) conditions. The difference between the matched and unmatched personalised conditions was not significant (p = 0.073, [−1.06, 0.03]).

In terms of the intention to change sugar consumption, ANCOVAs were conducted to compare intention to change across all conditions whilst controlling for sugar intake. Sugar intake score was calculated as the total score derived from all three items asking about sugar intake. Overall, there was a significant difference across condition in terms of intention to change sugar intake ($F(2, 208)$ = 9.53, p < 0.001). Significant differences were observed between the control and matched personalised conditions (p < 0.001, [1.98, 0.57]) and between matched and unmatched personalised conditions (p = 0.03, [0.05, 1.26]) but not between the control and unmatched personalised conditions (p = 0.069, [0.03, 1.27]).

Finally, in terms of intention to change consumption of saturated fat, analysis of covariance controlling for intake of saturated fat (as a total score of the four items asking about saturated fat intake) revealed an overall significant difference between conditions ($F(2, 215)$ = 13.40, p < 0.001). Post-hoc tests revealed a significant difference between the control condition and both unmatched and matched personalised conditions, p < 0.001 [−1.56, −0.39] and p < 0.001 [−1.96, −0.68] respectively. The difference between the unmatched and matched personalised conditions was not statistically significant (p = 0.403, [0.21, 0.91]).

4.3. Discussion

The current study aimed to understand whether food advice personalised according to socio-demographic, cognitive and sensory characteristics was more effective at motivating intention to make dietary changes in comparison to generic dietary advice.

The results partially supported our hypotheses, providing initial evidence for the effectiveness of addressing variables beyond genetic factors within personalised dietary advice Food choice advice matched to participants' cluster characteristics motivated intention to change all dietary aspects when compared to the generic advice. However, for vegetables and saturated fat, participants were also motivated when advice was not matched to their cluster. This could be because overall participants from these two conditions did not differ in the how personalised they perceived the message to be. An explanation for this lack of difference might be due to the presence of some overlap in cluster characteristics. For example, low household income is a characteristic of cluster 6 as well as cluster 25. Participants in cluster 25 that read a message designed with cluster 6 in mind, would have been classed as receiving a personalised unmatched message. For these individuals however, parts of the message, addressing how healthy ingredients can be cheap could still have felt personalised. Additionally, the lack of difference between matched and unmatched conditions regarding change in vegetables and saturated fat intake could be because, as opposed to generic advice, both matched and unmatched advice offered examples of vegetables and recommended alternatives to replace frequently consumed meals known to be high in saturated fat. Thus, providing examples and alternatives might be sufficient to motivate dietary intention to change. Currently, recommendation systems that align with a customer's flavour have not been systematically compared
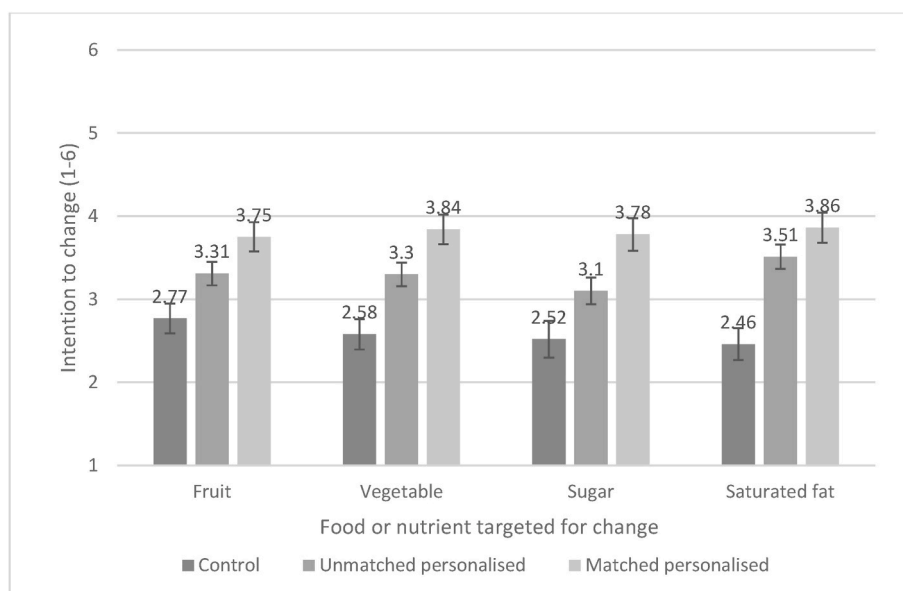

Fig. 5. Adjusted means and standard errors for intention to change intake each targeted food or nutrient across the three conditions.

to control conditions (recipes that do not align with consumer taste preferences) (e.g. Nag et al., 2017), therefore future work should seek to investigate this further.

However, the value of providing alternatives that suit participants taste preference became apparent when looking at the response concerning change in sugar intake. Although overall, participants reported a fairly low liking for alternatives (most likely because this was anticipated rather than based on actual tasting), when the message was matched to the cluster characteristics there was more of a sensory liking for the sugar alternative. The strong liking for the sugar alternative, of the participants who had advice matched to their cluster, is potentially why these participants were more strongly motivated to change their sugar intake when compared to both those who read generic or unmatched advice. This is also supported by the fact that overall, participants' liking for the recommended alternatives did relate to how personalised they felt the message was. These results provide evidence for the effect of personalised advice based on sensory characteristics and are in line with other evidence showing that consumers are more likely to select a healthy recipe if the recipe matches their preferences (Pecune, Callebert, & Marsella, 2020). In terms of fruit intake, the only significant difference in intention to change was observed between the control condition and the matched condition. Therefore, participants reading advice that was not matched to their cluster did not intend to change their fruit intake significantly more than the control condition, or significantly less than the matched condition. This could be due to the way messages were framed in relation to the cognitive variables. Indeed, fruit consumption messages were mostly framed to address the two cognitive factors. For example, discounting was addressed by stating that consuming fruit can immediately make one feel refreshed and energized. Similarly, the tendency to engage in heuristics was addressed by recommending increasing fruit intake by buying mixed bags of fruit for smoothies. These results are in line with evidence suggesting addressing cognitive biases can encourage healthy eating (Satia et al., 2010), and highlight the importance of tailoring advice in line with participants' cognitive factors, as only messages tailored to individual's cognitive biases can be strong enough to elicit an intention to change. However, due to the methodology used we cannot precisely identify whether the manipulation of certain factors was relevant for particular parts of the advice. Future studies should seek to address this by building advice with different levels of personalisation for each of the intended changes.

This study is not without limitations. First, we only assessed participants' intention to change rather than actual behaviour change. Whilst behavioural intentions have been found to predict behaviour change (Webb & Sheeran, 2006), future work should seek to understand whether advice personalised on several individual characteristics leads to actual behavioural change. Further to this point, due to the use of self-report measures, participant's responses could have been subject to desirability bias. To address these issue future research should consider a longitudinal study design using more objective measures of diet intake. Second, our advice was not able to consider participants' current dietary intake, namely we could not advise participants individually to either increase, maintain or decrease consumption of certain foods. Only a small proportion of participants did mention they were already following the dietary guidelines, as seen in response to the open-ended questions, and we did control for intake in our analyses. However, intake was assessed using only a few food-frequency items that can be subject to reporting bias and are unlikely to give an accurate picture of participants' total diet. This is potentially why, for most participants, habitual intake was not related to how suitable they thought the advice to alter their consumption was.

5. General discussion

Overall, our findings support previous research regarding the effective nature of personalised food choice advice as compared to generic

guidance (Ordovas et al., 2018). Additionally, our paper suggests a way to build advice based on a combination of individual characteristics and shows that this can be effective in motivating intention to change. Given the final study was conducted online, and the cluster modelling could be easy integrated into an online application, the findings also contribute to evidence regarding the feasibility of using diet apps to deliver effective food advice to the general population (Fallaize, Franco, Pasang, Hwang, & Lovegrove, 2019). Our approach also shows how to utilize existing datasets which contain only subsets of relevant variables for such interventions by creating a synthetic data set, therefore further highlighting how synthetic datasets can overcome limitations in data acquisition (Chan, 2024), and enable data-drive innovation within healthcare (Omotunde & Mouhamed, 2023).

Results assessing the effectiveness of the tool are particularly interesting given that the manipulation we attempted was minimal. Namely, our aim was to use generic advice as a template and only slightly change this according to the characteristics of the selected clusters. However, there are multiple other ways in which the characteristics we focused on could be addressed in order to create change. For example, pictures can increase attention and adherence to health information (Houts, Doak, Doak, & Loscalzo, 2006) and additionally viewing pictures of food can evoke retrieval of information about the taste of food (Avery, Liu, Ingeholm, Gotts, & Martin, 2021). Therefore, future work should seek to understand whether adding images of the recommended alternatives could potentially motivate participants even further.

A limitation of this study was that we could not separately differentiate the value of tailoring advice to each of the variables we targeted, as our advice was designed to address these factors concomitantly. Even though there is evidence to suggest the value of tailoring advice according to these variables, future research should seek to understand the added value of tailoring the advice to these variables concomitantly versus independently. Second, the cluster analysis was conducted on a synthetic data set, the creation of which is not without limitations. For example, any bias present in the original datasets could be unintentionally amplified during the creation of a synthetic dataset (Giuffrè & Shung, 2023). However, one of the strengths of our study is the use of the LCFS as a primary dataset, one of the most representative datasets in the UK, that employs multiple quality assurance measures to ensure that the LCFS data are as reliable as possible (Office for National Statistics, 2019), which would have mediated this risk. Furthermore, our synthetic dataset is not being used to generate broad conclusions that apply to the general population, but instead to provide personalised advice. In conclusion, our study finds some evidence that using existing datasets to segment the target population and subsequently tailor food choice advice according to multiple characteristics can be more effective in motivating dietary intention to change than generic advice. Future work should evaluate whether such advice leads to actual behavioural change over an extended period of time. If effective online materials could be incorporated into existing healthcare and advice apps, such as the NHS app, this could allow a more targeted nutritional advice strategy for all.

Open practices statement

All data are publicly available through the Open Science Framework at the following URL address: https://osf.io/yt3x DOI 10.17605/OSF.IO/YT3XD.

Ethical statement

Study 1a used pre-existing data sets where we confirm the University of Reading data sets were collected from studies that ran in accordance

with the Declaration of Helsinki and were given a favourable opinion for conduct by the University of Reading Research Ethics Committee (approval codes: UREC 15/37). Study 2, the intervention study, was similarly run in accordance with the Declaration of Helsinki and weas given a favourable opinion for conduct by the University of Reading Research Ethics Committee (approval codes: UREC 19/49). All participants provided informed consent.

CRediT authorship contribution statement

R.A. Briazu: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **L. Bell:** Writing – review & editing, Methodology, Investigation. **G.F. Dodd:** Writing – review & editing, Investigation. **S. Blackburn:** Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **C. Massri:** Writing – review & editing, Funding acquisition, Conceptualization. **B. Chang:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **S. Fischaber:** Software, Methodology, Funding acquisition, Formal analysis, Conceptualization. **A. Kehlbacher:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **C.M. Williams:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization. **L. Methven:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **R. McCloy:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

None.

Data availability

Link to data is provided in the manuscript

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appet.2024.107600.

References

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., … Murray, C. J. (2019). Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the global burden of disease study 2017. *The Lancet, 393*(10184), 1958–1972.

Amiri, M., Li, J., & Hasan, W. (2023). Personalized flexible meal planning for individuals with diet-related health concerns: System design and feasibility validation study. *JMIR Formative Research, 7*, Article e46434.

Appelhans, B. M., Tangney, C. C., French, S. A., Crane, M. M., & Wang, Y. (2019). Delay discounting and household food purchasing decisions: The SHoPPER study. *Health Psychology, 38*(4), 334.

Appelhans, B. M., Waring, M. E., Schneider, K. L., Pagoto, S. L., DeBiasse, M. A., Whited, M. C., et al. (2012). Delay discounting and intake of ready-to-eat and away-from-home foods in overweight and obese women. *Appetite, 59*(2), 576–584.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*(1), 124–140.

Avery, J. A., Liu, A. G., Ingeholm, J. E., Gotts, S. J., & Martin, A. (2021). Viewing images of foods evokes taste quality-specific activity in gustatory insular cortex. *Proceedings of the National Academy of Sciences, 118*(2), Article e2010932118.

Bakre, S., Shea, B., Ortega, K., Scharen, J., Langheier, J., & Hu, E. (2022). Changes in food insecurity among individuals using a telehealth and nutrition platform: Longitudinal study. *JMIR Formative Research, 6*(10), Article e41418.

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology and Health, 13*(4), 623–649.

Briazu, R. A., Masoud, F., Hunt, F., Pettinger, C., Wagstaff, C., & McCloy, R. (2024). Barriers and facilitators to healthy eating in disadvantaged adults living in the UK: A scoping review. *BMC Public Health, 24*, 1770. https://doi.org/10.1186/s12889-02 4-19259-2.

Burton, D. G. A., Wilmot, C., & Griffiths, H. R. (2018). Personalising nutrition for older adults: The InCluSilver project. *Nutrition Bulletin, 43*(4), 442–455.

Buttriss, J. L. (2016). The eatwell guide refreshed. *Nutrition Bulletin, 41*, 135–141.

Caswell, J. A., & Yaktine, A. L. (2013). *Committee on examination of the adequacy of food resources and SNAP allotments. Food and nutrition board. Committee on national staistics; institute of medicine.* Washington (DC): National Research Council. National Amademies Press (US).

Celis-Morales, C., Livingstone, K. M., Marsaux, C. F., Macready, A. L., Fallaize, R., Mathers, C. B., et al. (2017). Effect of personalized nutrition on health-related behaviour change: Evidence from the Food4Me European randomized controlled trial. *International Journal of Epidemiology, 46*, 578–588. https://doi.org/10.1093/ ije/dyw186

Chan, C. Y. (2024). *Data-driven innovation: The potential of synthetic data through generative AI.*

Chen, P. J., & Antonelli, M. (2020). Conceptual models of food choice: Influential factors related to foods, individual differences, and society. *Foods, 9*(12), 1898.

Clegg, M. E., & Williams, E. A. (2018). Optimizing nutrition in older people. *Maturitas, 112*, 34–38.

Cobiac, L. J., Scarborough, P., Kaur, A., & Rayner, M. (2016). The eatwell guide: Modelling the health implications of incorporating new sugar and fibre guidelines. *PLoS One, 11*(12), Article e0167859.

D'Orazio, M. (2019). StatMatch: Statistical matching or data fusion. Retrieved from https://CRAN.R-project.org/package=StatMatc.

d'Angelo, C., Guthrie, S., Draper, A., & Gloinson, E. R. (2020). *Food consumption in the UK: Trends, attitudes and drivers.* Europe: RAND.

Das, H. P., Tran, R., Singh, J., Yue, X., Tison, G., Sangiovanni-Vincentelli, A., et al. (2022). Conditional synthetic data generation for robust machine learning applications with limited pandemic data. *Proceedings of the AAAI Conference on Artificial Intelligence, 36*(11), 11792–11800.

Davis, P., Lay-Yee, R., & Pearson, J. (2010). Using micro-simulation to create a synthesised data set and test policy options: The case of health service effects under demographic ageing. *Health Policy, 97*, 267–274.

Deglaire, A., Méjean, C., Castetbon, K., Kesse-Guyot, E., Hercberg, S., & Schlich, P. (2015). Associations between weight status and liking scores for sweet, salt and fat according to the gender in adults (The Nutrinet-Sante study). *European Journal of Clinical Nutrition, 69*(1), 40–46.

European Commission. (2013). *Statistical matching: A model based approach for data integration.* Luxembourg: Publications Office of the European Union. https://doi.org/10.2785/44822. Commission, European.

Fallaize, R., Franco, R. Z., Pasang, J., Hwang, F., & Lovegrove, J. A. (2019). Popular nutrition-related mobile apps: An agreement assessment against a UK reference method. *JMIR mHealth and uHealth, 7*(2), Article e9838.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Fosdick, B. K., DeYoreo, M., & Reiter, J. P. (2016). Categorical data fusion using auxiliary information. *Annals of Applied Statistics, 10*(4), 1907–1929. https://doi.org/10.1214/16-AOAS925

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25–42.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(2), 351–401.

Giuffrè, M., & Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digital Medicine, 6*(1), 186.

Headey, D. D., & Alderman, H. H. (2019). The relative caloric prices of healthy and unhealthy foods differ systematically across income levels and continents. *The Journal of nutrition, 149*(11), 2020–2033.

Heinz, D. (2015). Nonparametric mixed membership models. In D. M. E. M. Airoldi (Ed.), *Handbook of mixed memeberhip models and their applications* (pp. 89–116). CRC Press, Taylor & Francis Group.

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling, 61*(2), 173–190.

Hu, J., Reiter, J. P., & Wang, Q. (2018). Dirichlet process mixture models for modeling and generating synthetic versions of nested categorical data. *Bayesian Analysis, 13*(1), 183–200.

Islam, K. F., Awal, A., Mazumder, H., Munni, U. R., Majumder, K., Afroz, K., … Hossain, M. M. (2023). Social cognitive theory-based health promotion in primary care practice: A scoping review. *Heliyon, 9*, Article e14889.

Jabs, J., & Devine, C. M. (2006). Time scarcity and food choices: An overview. *Appetite, 47*(2), 196–204.

Jarmolowicz, D. P., Bickel, W. K., Sofis, M. J., Hatz, L. E., & Mueller, E. T. (2016). Sunk costs, psychological symptomology, and help seeking. *SpringerPlus, 5*, 1–7.

Jinnette, R., Narita, A., Manning, B., McNaughton, S. A., Mathers, J. C., & Livingstone, K. M. (2021). Does personalized nutrition advice improve dietary intake in healthy adults? A systematic review of randomized controlled trials. *Advances in Nutrition, 12*(3), 657–669.

Jones, N. R., Tong, T. Y., & Monsivais, P. (2018). Meeting UK dietary recommendations is associated with higher estimated consumer food costs: An analysis using the national

diet and nutrition survey and consumer expenditure data, 2008–2012. *Public Health Nutrition, 21*(5), 948–956.

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., … Weller, A. (2022). *Synthetic Data–what, why and how?*. arXiv preprint arXiv:2205.03257.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341.

Kurth-Nelson, Z., Bickel, W., & Redish, A. D. (2012). A theoretical account of cognitive effects in delay discounting. *European Journal of Neuroscience, 35*(7), 1052–1064.

Leitch, M. A., Morgan, M. J., & Yeomans, M. R. (2013). Different subtypes of impulsivity differentiate uncontrolled eating and dietary restraint. *Appetite, 69*, 54–63.

Liem, D. G., & Russell, C. G. (2019). The influence of taste liking on the consumption of nutrient rich and nutrient poor foods. *Frontiers in Nutrition, 174*.

Liverani, S., Hastie, D. I., Azizi, L., Papathomas, M., & Richardson, S. (2015). PReMiuM: An R package for profile regression mixture models using dirichlet processes. *Journal of Statistical Software, 64*(7), 1–30. Retrieved from http://www.jstatsoft.org/v64/i07/.

Mathers, J. C. (2019). Paving the way to better population health through personalised nutrition. *EFSA Journal, 17*, Article e170713.

Milkman, K. L., & Beshears, J. (2009). Mental accounting and small windfalls: Evidence from an online grocer. *Journal of Economic Behavior & Organization, 71*(2), 384–394.

Miller, S. R., & Knudson, W. A. (2014). Nutrition and cost comparisons of select canned, frozen, and fresh fruits and vegetables. *American Journal of Lifestyle Medicine, 8*(6), 430–437.

Nag, N., Pandey, V., & Jain, R. (2017). Live personalized nutrition recommendation engine. In *Proceedings of the 2nd international workshop on multimedia for personal health and health care* (pp. 61–68).

Office for National Statistics. (2017). Department for Environment, Food and Rural Affairs. In *Living Costs and Food Survey, 2015-2016. [data collection]* (2nd Release.). UK Data Service.

Office for National Statistics. (2019). *Living costs and food survey* (3rd Release). UK Data Service. SN, Article 2000028. https://doi.org/10.5255/UKDA-Series-2000028 [data series].

Omotunde, H., & Mouhamed, M. R. (2023). The modern impact of artificial intelligence systems in healthcare: A concise analysis. *Mesopotamian Journal of Artificial Intelligence in Healthcare, 2023*, 66–70.

Ordovas, J. M., Ferguson, L. R., Tai, E. S., & Mathers, J. C. (2018). Personalised nutrition and health. *BMJ, 361*.

Pecune, F., Callebert, L., & Marsella, S. (2020). A recommender system for healthy and personalized recipes recommendations. In *HealthRecSys@ RecSys* (pp. 15–20).

Pérez-Troncoso, D., Epstein, D. M., & Castañeda-García, J. A. (2021). Consumers' preferences and willingness to pay for personalised nutrition. *Applied Health Economics and Health Policy, 19*(5), 757–767.

Pliner, P., & Hobden, K. (1992). Development of a scale to measure the trait of food neophobia in humans. *Appetite, 19*(2), 105–120.

Roininen, K., Lähteenmäki, L., & Tuorila, H. (1999). Quantification of consumer attitudes to health and hedonic characteristics of foods. *Appetite, 33*(1), 71–88.

Satia, J. A., Barlow, J., Armstrong-Brown, J., & Watters, J. L. (2010). Qualitative study to explore Prospect Theory and message framing and diet and cancer prevention-related issues among African American adolescents. *Cancer Nursing, 33*(2), 102–109.

Schulte-Mecklenbeck, M., Sohn, M., de Bellis, E., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite, 71*, 242–251.

Scott, C., Sutherland, J., & Taylor, A. (2018). *Affordability of the UK's Eatwell guide*. The Food Foundation.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 639–650.

Smithers, G. (1993). MAFF'S nutrient databank. *Nutrition & Food Science, 93*(2), 16–19. https://doi.org/10.1108/eb059385

Steenson, S., & Buttriss, J. L. (2021). Healthier and more sustainable diets: What changes are needed in high-income countries? *Nutrition Bulletin, 46*(3), 279–309. https://doi.org/10.1111/nbu.12518

Suher, J., Raghunathan, R., & Hoyer, W. D. (2016). Eating healthy or feeling empty? How the "healthy= less filling" intuition influences satiety. *Journal of the Association for Consumer Research, 1*(1), 26–40.

Turnwald, B. P., & Crum, A. J. (2019). Smart food policy for healthy food labeling: Leading with taste, not healthiness, to shift consumption and enjoyment of healthy foods. *Preventive Medicine, 119*, 7–13.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science, 185*(4157), 1124–1131.

Wade, S. (2015). *mclust.ext: Point estimation and credible balls for Bayesian cluster analysis*.

Wade, S., & Ghahramani, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis, 13*(2), 559–626.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132*(2), 249.