Knowledge Representations in Health Judgments

Natasha Gandhi (N.Gandhi.1@warwick.ac.uk)
Applied Psychology, WMG, University of Warwick
Coventry, CV4 7AL, United Kingdom

Wanling Zou (wanlingz@sas.upenn.edu)
Department of Psychology, University of Pennsylvania
Philadelphia, PA 19104 USA

Caroline Meyer
Applied Psychology, WMG
University of Warwick
Coventry, CV4 7AL, United Kingdom

Sudeep Bhatia
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104 USA

Lukasz Walasek
Department of Psychology
University of Warwick
Coventry, CV4 7AL, United Kingdom

Abstract

In the present paper, we introduce a novel computational approach for uncovering mental representations underlying healthiness judgments for food items. Using semantic vector representations derived from large-scale natural language data, we quantify the complex representations that people hold about foods, and use these representations to predict how both lay decision makers and experts (trained dietitians) judge the healthiness of food items. We also successfully predict the impact of behavioral interventions (e.g. the provision of nutrient content information or "traffic-light labels") on healthiness judgments for food items. Our models are highly general, and are capable of making predictions for nearly any food item. Finally, these models outperform competing models based on factual nutritional content, suggesting that health judgments depend more on complex (semantic) knowledge representations than on quantified nutritional information. The results in this paper illustrate how methods from cognitive science and computational linguistics can be combined with existing theories in psychology, to better predict, understand, and influence health behavior.

Keywords: judgment; knowledge representations; vector semantics; behavioral interventions; computational models

Introduction

Is granola healthy? What about steak? People make healthiness judgments daily and understanding the psychological underpinnings of these judgments is central to the development of effective health interventions. It is commonly believed that people's judgments of food healthiness are closely tied to their beliefs about the food's nutritional content. There is now a large body of research investigating whether people's judgments of food healthiness can be explained by government-provided nutritional guidelines (Bucher, Müller, & Siegrist, 2015; Rizk & Treat, 2014). In a typical study of this kind, participants are asked to rate the healthiness of some food stimuli and then explain their ratings. However, as self-assessed knowledge is not always reliable, it is difficult for participants to accurately state the rationale behind their beliefs and judgments (Fernbach, Light, Scott, Inbar, & Rozin, 2019; Schwarz & Clore, 1983). In addition, the nutritional content of foods does not often align perfectly with people's judgments (Rozin, 1996; Rozin, Fischler, Imada, Sarubin, & Wrzesniewski, 1999), and beliefs about food healthiness could stem from other complex associations that are uncorrelated with nutrient content. Thus the knowledge representations that underpin judgment may be biased, causing systematic (and thus predictable) error in healthiness judgments. In fact, the knowledge representations that people have for food items are often a product of social communication, media, and advertisements (Paquette, 2005; Provencher & Jacob, 2016; Yarar & Orth, 2018), which are sometimes at odds with nutritional guidelines (e.g. involve misleading claims such as "fat-free", "organic" and "no added sugars") (André, Chandon, & Haws, 2019; Steinhauser & Hamm, 2018).

Existing literature offers little insights into the extent to which these non-nutrient-related media and informational factors influence judgments of healthiness, as they are difficult to identify and measure. Additionally, despite extensive research on the effects of different formats of nutrient labelling and the provision of nutrient information (Cecchini & Warin, 2016), a single coherent and evidence-based labelling strategy is still to be determined (Goiana-da Silva et al., 2019). Part of the problem is that the effectiveness of nutrient labelling systems is dependent on the existing knowledge, beliefs, and associations about that food item (Ikonen, Sotgiu, Aydinli, & Verlegh, 2019). As knowledge representations for food items are hard to measure, current evaluative approaches cannot make conclusive generalisations about the effectiveness of various nutrient-labelling systems and related public health interventions beyond the food stimuli used in particular studies.

In order to predict healthiness judgments of everyday food items, we thus need to model the complex representations and associations that people have for food items; representations that stem not from nutrient labelling but rather from the rich (and sometimes misleading) information presented in various forms of media. Fortunately, there have been recent advances in computational linguistics that offer a solution to this problem. These advances rely on the structure of word distribution in large-scale natural language data to uncover quantitative knowledge representations for words and phrases (Landauer & Dumais, 1997; see Lenci, 2018; Jones, Willits, Dennis, & Jones, 2015 for a review), such as those that describe natural entities like food items. These representations often take the form of high-dimensional semantic vectors for words (also known as word embeddings). The proximities between these vectors measure the associations between words, which in turn correlate with human semantic judgment, factual judgment, probability judgment, and social judgment (Bhatia, 2017a, 2017b; Bhatia & Walasek, 2019;

Caliskan, Bryson, & Narayanan, 2017; Garg, Schiebinger, Jurafsky, & Zou, 2018; Hills, Jones, & Todd, 2012; Mandera, Keuleers, & Brysbaert, 2017; Pereira, Gershman, Ritter, & Botvinick, 2016), all of which rely on association as a psychological cue. As the semantic vectors quantify what people know about various natural entities, they have been used to approximate knowledge representations of these entities and to predict more complex judgments, such as risk perception, consumer judgment, and organizational judgment (Bhatia, 2019; Richie, Zou, & Bhatia, 2019).

In the present study, we collect healthiness judgments for 172 food items from general public and registered dietitians and assess the effectiveness of three different behavioral interventions – calorie labelling, monochrome front of package (FoP) labelling, and traffic light (TL) colored FoP labelling. Using semantic vectors as knowledge representations of food items, we build a computational model to predict, in an out-of-sample manner, healthiness judgments for foods in different populations, as well as differences in such judgments between individuals exposed to different behavioral interventions. Finally, unlike previous approaches, there is no reliance on nutritional content values of the food stimuli. Therefore, this approach can be applied even to food items for which nutritional information is unknown. In the following pages, we illustrate the generalizability, accuracy, and power of our approach.

Methods

Participants

Participants from all but one study (1B) were recruited from Prolific Academic `https://www.prolific.ac`, an online crowdsourcing site designed for experimental research recruitment (Palan & Schitter, 2018). These participants were all from the general population. For Study 1B which required an expert sample, registered dietitians were contacted either by email or through social media sites to complete the online study. All participants were over the age of 18 and English-speaking, with no other constraints to the eligibility criteria in Study 1A-3. In Study 4, only UK residents were recruited because of their familiarity with the traffic light (TL) nutrient labelling presentation. Each participant was only eligible to take part in one of our studies. The overall target sample size was approximately 700 participants, and was determined before obtaining the data. This target sample size was chosen based on previous work, as this study adopts parts of the methodology and data analysis of the research by Bhatia (2019). The participants took part in return of a payment that equated to roughly £5.00/h, in line with the fair pay agreements of Prolific Academic. This research was approved by the University of Warwick's Biomedical and Scientific Research Ethics Sub-Committee (approval REGO-2018-2268).

There were 134 participants (mean age = 30.25 years, SD = 8.86, 43% females, and 84% had no dietary restrictions) in Study 1A and 19 registered dietitians (mean age = 37 years, SD = 10.36, 89% females and 68% had no dietary restrictions) in Study 1B. There were 197 participants (mean age = 30.30 years, SD = 10.74, 52% female, and 80% had no dietary restrictions) in Study 2, 195 participants (mean age = 29.16 years, SD = 10.28, 48% female, and 82% had no dietary restrictions) in Study 3, and 202 participants (mean age = 34.69 years, SD = 11.51, 70% female, and 81% had no dietary restrictions) in Study 4.

Table 1: Presentation formats of food names for each study and each condition.

Study	Control Condition	Experimental Condition
1A	Food Name Only	n/a
1B	Food Name Only	n/a
2	Food Name Only	Food Name + Calorie Content
3	Food Name Only	Food Name + Calories Content + FoP Content
4	Food Name Only	Food Name + Calories Content + FoP Content + TL labeling

Stimuli

The initial list of foods was taken from the USDA Food Composition Database, the most recent official publication of nutrient information pertaining to over 3102 unique food items (USDA, 2018). Only foods present in the pre-trained word embedding model were considered, leaving a subset of 571 food items. Two hundred food items, across all food categories (e.g. vegetables, meats, dishes), were then manually chosen to maximise variance of the calorie values. Unknown and ambiguous food items were also removed through double blind coding, resulting in the final list of 172 usable food items. The presentation format of the key nutrient information in the experimental conditions of Study 2-4 was based on guidance from UK government publications (Department of Health and Social Care, 2013).

Design and Procedure

A between-subjects design was used to explore how displaying nutrient information, akin to existing policy interventions, influences people's representations of food healthiness. The sub-studies (see Table 1) were all conducted between December 2018 and April 2019, with all recruitment per sub-study completed on the same day. After providing consent (Study 1A and 1B), and being randomly assigned to a condition (Study 2-4), participants were instructed to rate the healthiness of all 172 food items. The scale ranged from -100 (extremely unhealthy) to +100 (extremely healthy); the starting slider position was always defaulted at zero (neither healthy nor unhealthy). This scale was chosen because it is fine-grained (200 intervals) and balanced (symmetric around

0), as well as being consistent with previous relevant studies (Bhatia, 2019; Rizk & Treat, 2014). Participants also had the option of selecting “Don’t know” if they were unfamiliar with a food item. The order of the items was randomised for every participant and only one item was visible at a time. The same generic task instruction: “Using the slider, please use your first impression to rate the following food item according to the scale below:” was displayed above all stimuli in every study condition. Information asking about participants’ birth year, gender and dietary restrictions was collected at the end of the study, as well as years of experience as a registered dietitian and area of specialism for our dietitian sample.

Computational Approach

In all studies, we used three statistical models to predict subjective food healthiness judgments. Our analysis explored participant judgements at the aggregate level, averaging food item ratings within each condition of every study. We evaluated the accuracy of each of our three statistical models in predicting subjective food healthiness judgments using leave-one-out cross validation [1], which means that we trained our models on all but one participant-supplied judgement and used the trained model to predict the rating of the left-out food item. We repeated this procedure for all food items. This ensured that our modelling avoided overfitting and that performance of each model was evaluated based on model generalisability.

Our first model was the nutrient model, in which we used nutrient content information to predict healthiness judgements. Using OLS regression, we predicted ratings using the following nutrients: food calorie content, amounts of nutrients (fat, saturates, sugar, salt and protein) per 100g, and traffic light color coding (green, orange and red).

In the vector representation model, we used vector representations from the word2vec model to approximate the general knowledge people associate with the healthiness of our food stimuli names. Our model is pre-trained on a dataset of Google News articles, which has 300-dimensional vector representations for the three million most common words and phrases in the English language (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) [2]. In our main analyses, we used normalized word2vec vectors, which represent the most commonly associated words with our food stimuli, to predict the participant supplied healthiness ratings. Because of the high number of predictor variables in this model (300), we applied a regularised regression technique known as ridge regression. Ridge regression allows high numbers of predictors to be considered and takes into account whether predictors are highly correlated. In the previous and similar work (e.g. Bhatia, 2019; Richie et al., 2019), ridge regression was found to be the best-fitting regression technique for mapping pre-trained 300-dimensional vector representations to judgements and was consequently chosen for our analysis [3].

Finally, our third and final model combines the vector representation model and the nutrient model, also using ridge regression to explore the extent that both models can collectively explain people’s subjective food healthiness judgements.

Results

We begin by showing the distribution of aggregate healthiness ratings from Study 1A in Figure 1. Here we can see that healthiness judgements vary greatly amongst the food stimuli, both across and within food categories. Unsurprisingly, the foods with the healthiest ratings were all fruit and vegetables, with the top five mean ratings ranging between 82 – 77 for broccoli, carrots, apple, cucumber and tomatoes respectively. The five foods that received the unhealthiest ratings, ranging between -65 and -50, were cola, donuts, skittles, cheeseburger, and kit kat. A sample of the 172 food stimuli can also be seen in Figure 1, highlighting the variety of foods used to train our computational models.

We now turn to our main analysis, in which we attempted to predict the aggregate judgments of healthiness using nutrient model, vector representation model, or the combined model. Figure 2 summarises the out-of-sample coefficient of determination (r^2) of these three models, separately for each condition across all five studies. The dots within each scatterplot represent the predicted vs. actual (aggregated) healthiness ratings for the foods. R^2 was calculated as the squared pearson correlation between actual ratings and predicted ratings by the models using leave-one-out cross validation.

As shown in Figure 2, the vector representation model performed very well across all studies and conditions. In fact, the predictive accuracy of the vector representation model was consistently between 76-77% in the control conditions of all four studies. Predictive accuracy for this model was slightly lower in our expert sample, perhaps because they relied more on internal nutrient knowledge about the foods.

We can also assess how different types of nutrient labelling affects the predictive ability of the model using word vectors (experimental conditions of Studies 2-4). For example, despite participants being provided with calorie content information in Study 2, word vectors were equally predictive of healthiness judgments compared to when participants were only provided with food names. However, we can start to see a reduction in reliance on associations, as captured by our word vector model, when participants were provided with monochrome nutrient labelling, and particularly traffic light labelling.

[1] We also tried cross validation with other train-test splits, e.g.,9-1. The results were similar. We will publish these results separately.

[2] We chose this particular pre-trained word embeddings due to its prior success in predicting various human judgments (Bhatia, 2019; Richie et al., 2019). We also fit this model with vector representations from other pre-trained word embeddings, such as fastText (Mikolov et al., 2018) and GloVe (Pennington, Socher, & Manning, 2014). Results will be published separately.

[3] We also tested other regression techniques including lasso, support vector, and k-nearest neighbors regression and found that ridge regression is indeed the best-fitting regression. Results will be published separately.

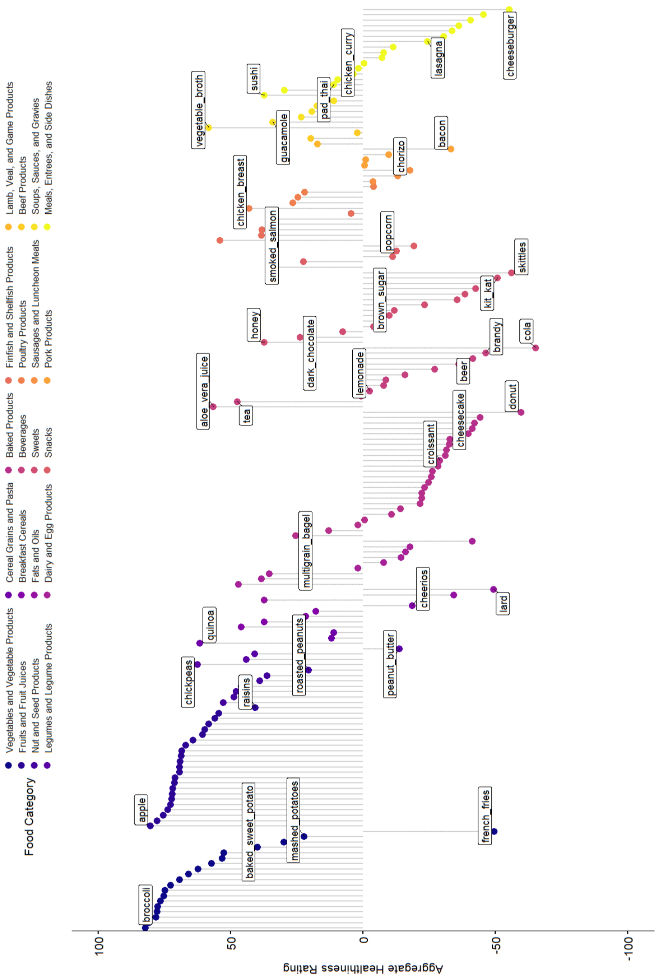

Figure 1: Distribution of aggregated food healthiness ratings from Study 1A. Foods have been separated by food category.

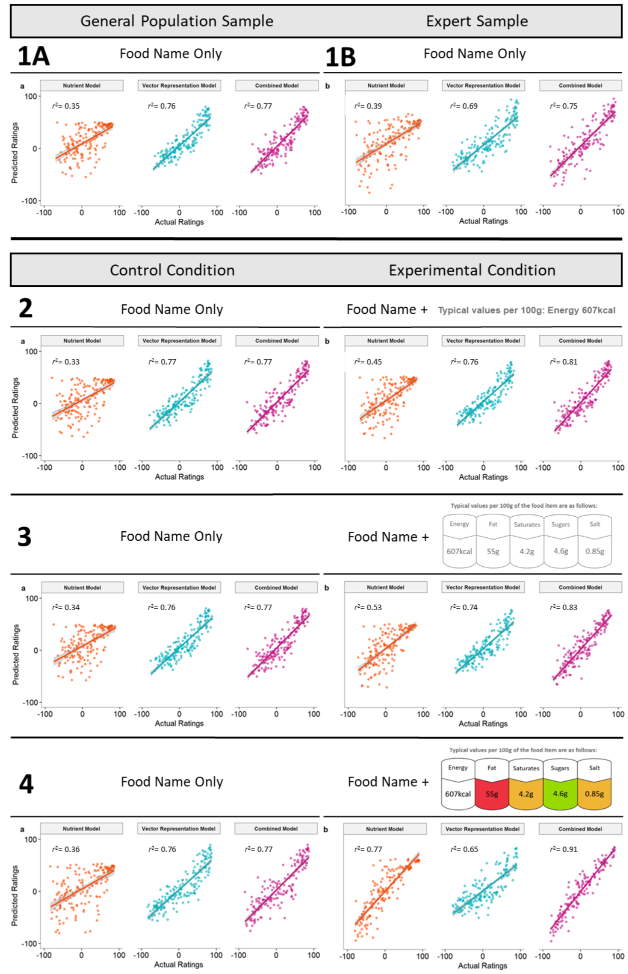

Figure 2: Leave-one-out cross validation results for the control and treatment conditions of each study, comparing models predicting health ratings of foods using only nutrient content, only semantic vector representations, or a combination of nutrient content and semantic vector representations. All conditions of Studies 2-4 were conducted in a general public sample.

By comparison, the predictive accuracy of the model based on the nutritional information varies greatly between studies and conditions. In all control conditions, the out-of-sample predictive accuracy of the nutrient model was considerably lower than the vector representation model. In line with expectations, the nutrient model was slightly better at predicting healthiness ratings of experts than of those from the general population sample. However, it is still clear to see that the out-of-sample predictive power of the vector representation model is much better than that of the nutrient model, even for experts. When people were presented with additional nutritional information about foods, the predictive performance of the nutrient models increased more substantially, suggesting that the nutrient labelling information presented to participants did influence their healthiness judgments to varying extents. The highest accuracy of the nutrient model was achieved for the ratings by participants who saw calories, nutrients and relative magnitudes (as indicated by the traffic light coloring scheme), where the predictive accuracy reached 77%. This is also the only group for which the accuracy was higher than for the corresponding results of vector representation model.

If we turn our attention to the findings of the combined model in Figure 2, comprising of both the vector representation model and the nutrients model, we can see that it achieves very high predictive accuracy. In fact, in the case of models fitted to ratings made by the participants who saw either monochrome labelling or traffic light colored labels, the predictive accuracy of the combined model even exceeds the accuracy of the vector representation model. These results support the interpretation that word vectors explain people's judgments over and above the nutritional information of individual foods. Of course, if nutritional content information of foods were always available then the best predictive model of people's subjective food healthiness judgments would be

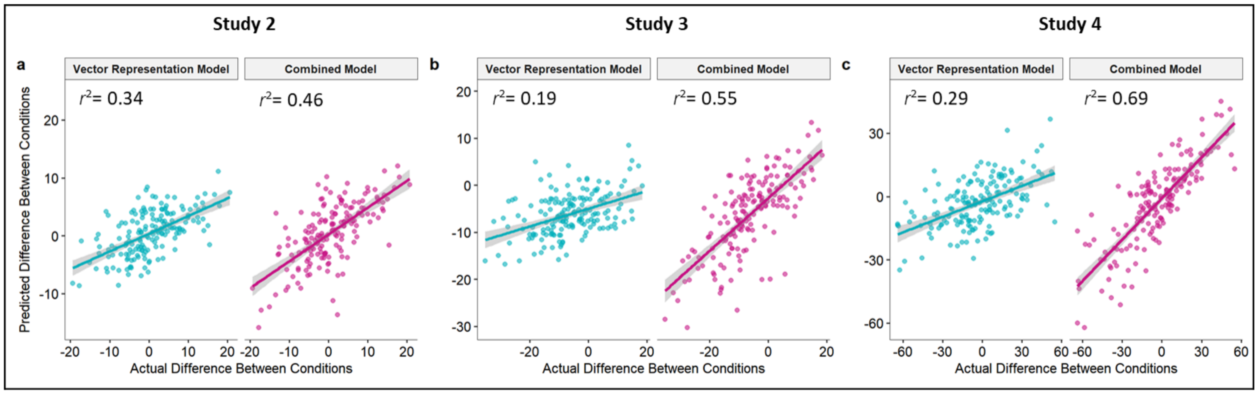

Figure 3: Leave-one-out cross validation results for the ability of vector representations and the combined model to predict the difference between conditions in Study 2 (Calorie – Control), Study 3 (Front of Package Labelling – Control) and Study 4 (Traffic Light Labelling – Control).

the combined model. Nonetheless, these findings demonstrate the high capability of the vector representation model in providing subjective healthiness insights even when just food names are known. We obtained similar findings on the individual-level analyses [4].

We are also able to assess how much of the variance between the control and experimental condition of each study can be explained by word vectors alone. We show this in Figure 3 alongside the combined model to demonstrate the maximum predictive power of our computational models in explaining what alters people's healthiness judgments. Associations, as captured by the word vector model, explain a significant proportion of the difference between those presented with just food names and additional calorie content information (seen in Panel a). The model using word vectors is less predictive of the changes between conditions in Study 3 and 4 but the predictive ability of the combined model considerably increased. This reinforces that participants were making use of the information from the respective monochrome and traffic light colored nutrient labelling formats to influence their judgments in these experimental conditions.

Another benefit of the vector representation approach is that it can identify regions of the semantic space related to food healthiness. This can be done by passing the vector representations of common words (that are not necessarily food items) through a model trained on participants' food healthiness judgments. Words given high predictions would be those most associated with healthiness, and would capture the conceptual underpinnings of health judgment. Figure 4 shows a word cloud of the fifty English language words with the highest healthiness predictions, derived with this approach. Visibly, agriculture and nature related words, such as *crop*, *organic*, and *leaf*, make up the majority of this word cloud. Interestingly, the word *healthy* is also present in the word cloud even though our model was never explicitly trained on this concept. It seems that implicit in people's judgments are associations with concepts like healthiness, as well as other concepts (e.g. naturalness, organic, appearance) identified by previous researches as being psychological cues for food healthiness. Our novel computational approach provides quantitative methods for uncovering these associations.

General Discussion

We combined insights from cognitive science and computational linguistics to uncover knowledge representations underlying health judgments and built a computational model based on such representations to predict people's subjective healthiness ratings. We showed that this model achieved high accuracy, with an out of sample predictive accuracy of up to 77% for 172 diverse foods. Notably, we found that semantic vector representations of foods were an even better predictor of health judgments than any internal knowledge that dietitians hold about the foods' nutritional values. This is in line with previous literature that found that, contrary to expectations, nutritional expertise does not always translate into higher reliance on nutritional information when making healthiness judgments (Orquin, 2014). In contrast to classical research that often assumes an existence of a direct mapping between nutritiousness and healthiness (Bucher, Hartmann, Rollo, & Collins, 2017), our results show that mere nutritional information may only be a fragment of mental representations of food items.

We were also able to interpret why the vector representation model performs well. Using our best-fit model on participant healthiness rating data to infer the associations implicit in people's judgment, we found that healthy food items were strongly associated with words related to nature and the cultivation of vegetarian food products (e.g., "crop", "harvest", and "agricultural"). This means that even in the presence of interventions aimed at aiding individuals to choose healthier options, internal knowledge representations continue to exert a strong effect on people's judgments. This is consistent with

[4]Results will be published separately.

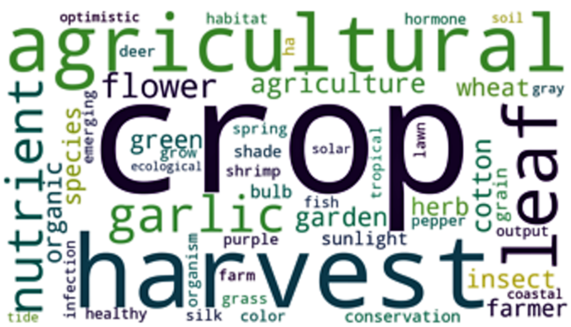

Figure 4: Words with the highest predicted healthiness rating based on the Vector Representation Model estimated for the lay people in Study 1A (control).

the findings showing that marketing techniques stressing naturalness and organic nature of food create a halo effects that can influence people's judgments of food healthiness (Schuldt & Schwarz, 2010; Scrinis & Parker, 2016; Whalen, Harrold, Child, Halford, & Boyland, 2018). In line with this research, our work also shows that nutritional labelling neither substitutes nor corrects for the associations that people rely on when judging food's healthiness. This is why even with nutrition labels it is possible to predict health judgment using vector representations of knowledge and association, with a high degree of accuracy.

Our approach offers a unique insight into the psychological basis of subjective food healthiness judgements. We obtained healthiness ratings by providing participants with the generic name of the food stimuli (with or without nutritional information) to gain universal insights into people's healthiness judgment for that food. Written language predominantly contains generic information about foods because its primary purpose is to communicate information to others (De Deyne, Perfors, & Navarro, 2016). Therefore, our vector representation model is most suitable, and indeed highly successful, at approximating mental representations of food items at the abstract level. However, a model trained on written text is unlikely to capture attributes of foods perceived to be uninformative, like specific visual attributes of the food (e.g. freshness and size). In fact, Richie et al. (2019) found that representations derived from the exact same word embeddings (i.e. word2vec, Mikolov et al., 2013) did not predict food tastiness judgments well. We also suspect that our vector representation model may not work as well for uncommon food items, as there may not be sufficient natural language data to derive tractable representations for food items that people rarely talk about. Hence, while our results are certainly promising, our findings are a first step in providing a rich set of attributes and associations that people use in judging food's healthiness. The direction of future studies should be to assess how the word vectors might perform when the judgments are influenced by additional sensory information, including promotional packaging, smell of food, portion sizes, or hunger states.

Using three design features from different nutrient labelling formats, we demonstrate how the interaction between external information and people's pre-existing knowledge about foods can be uncovered using computational models. Our findings are complementary to existing reviews about the effectiveness of interventions, and provide robust evidence of the differences in ability of various types of nutrient labelling to shift healthiness judgments (Egnell, Talati, Hercberg, Pettigrew, & Julia, 2018). Using our approach, we are able to assess and compare how each design feature of currently implemented labelling strategies alters people's judgements and reliance on knowledge representations of foods. This is important as it could potentially provide rationale for removing unnecessary information that could contribute to overload confusion (Leek, Szmigin, & Baker, 2015).

In summary, our modelling approach can be used by registered dietitians, health professionals, policy makers and researchers alike to gain better insights into subjective food judgments, which is pivotal to confronting the obesity epidemic. In doing so, our paper shows how established ideas and methods in cognitive science can be used to guide behavioral outcomes and improve welfare in subject populations.

References

André, Q., Chandon, P., & Haws, K. (2019). Healthy through presence or absence, nature or science?: A framework for understanding front-of-package food claims. *Journal of Public Policy & Marketing*, *38*(2), 172–191.

Bhatia, S. (2017a). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20.

Bhatia, S. (2017b). The semantic representation of prejudice and stereotypes. *Cognition*, *164*, 46–60.

Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, *65*(8), 3800–3823.

Bhatia, S., & Walasek, L. (2019). Association and response accuracy in the wild. *Memory & cognition*, *47*(2), 292–298.

Bucher, T., Hartmann, C., Rollo, M. E., & Collins, C. E. (2017). What is nutritious snack food? a comparison of expert and layperson assessments. *Nutrients*, *9*(8), 874.

Bucher, T., Müller, B., & Siegrist, M. (2015). What is healthy food? objective nutrient profile scores and subjective lay evaluations in comparison. *Appetite*, *95*, 408–414.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Cecchini, M., & Warin, L. (2016). Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies. *Obesity reviews*, *17*(3), 201–210.

De Deyne, S., Perfors, A., & Navarro, D. J. (2016). Predicting human similarity judgments with distributional models: The value of word associations. In *Proceedings of coling*

2016, the 26th international conference on computational linguistics: Technical papers (pp. 1861–1870).

Egnell, M., Talati, Z., Hercberg, S., Pettigrew, S., & Julia, C. (2018). Objective understanding of front-of-package nutrition labels: An international comparative experimental study across 12 countries. *Nutrients*, *10*(10), 1542.

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., & Rozin, P. (2019). Extreme opponents of genetically modified foods know the least but think they know the most. *Nature Human Behaviour*, *3*(3), 251–256.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644.

Goiana-da Silva, F., Cruz-e Silva, D., Miraldo, M., Calhau, C., Bento, A., Cruz, D., … Araújo, F. (2019). Front-of-pack labelling policies and the need for guidance. *The Lancet Public Health*, *4*(1), e15.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological review*, *119*(2), 431.

Ikonen, I., Sotgiu, F., Aydinli, A., & Verlegh, P. W. (2019). Consumer effects of front-of-package nutrition labeling: an interdisciplinary meta-analysis. *Journal of the Academy of Marketing Science*, 1–24.

Jones, M. N., Willits, J., Dennis, S., & Jones, M. (2015). Models of semantic memory. *Oxford handbook of mathematical and computational psychology*, 232–254.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211–240.

Leek, S., Szmigin, I., & Baker, E. (2015). Consumer confusion and front of pack (fop) nutritional labels. *Journal of Customer Behaviour*, *14*(1), 49–61.

Lenci, A. (2018). Distributional models of word meaning. *Annual review of Linguistics*, *4*, 151–171.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2018). Advances in pre-training distributed word representations. In *Proceedings of the international conference on language resources and evaluation (lrec 2018).*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Orquin, J. L. (2014). A brunswik lens model of consumer health judgments of packaged foods. *Journal of Consumer Behaviour*, *13*(4), 270–281.

Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Paquette, M.-C. (2005). Perceptions of healthy eating: state of knowledge and research gaps. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique*, S15–S19.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive neuropsychology*, *33*(3-4), 175–190.

Provencher, V., & Jacob, R. (2016). Impact of perceived healthiness of food on food choices and intake. *Current obesity reports*, *5*(1), 65–71.

Richie, R., Zou, W., & Bhatia, S. (2019). Predicting high-level human judgment across diverse behavioral domains. *Collabra: Psychology*, *5*(1), 50.

Rizk, M. T., & Treat, T. A. (2014). An indirect approach to the measurement of nutrient-specific perceptions of food healthiness. *Annals of Behavioral Medicine*, *48*(1), 17–25.

Rozin, P. (1996). The socio-cultural context of eating and food choice. In *Food choice, acceptance and consumption* (pp. 83–104). Springer.

Rozin, P., Fischler, C., Imada, S., Sarubin, A., & Wrzesniewski, A. (1999). Attitudes to food and the role of food in life in the usa, japan, flemish belgium and france: Possible implications for the diet–health debate. *Appetite*, *33*(2), 163–180.

Schuldt, J. P., & Schwarz, N. (2010). The" organic" path to obesity? organic claims influence calorie judgments and exercise recommendations. *Judgment and Decision making*, *5*(3), 144–150.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513–523.

Scrinis, G., & Parker, C. (2016). Front-of-pack food labeling and the politics of nutritional nudges. *Law & Policy*, *38*(3), 234–249.

Steinhauser, J., & Hamm, U. (2018). Consumer and product-specific characteristics influencing the effect of nutrition, health and risk reduction claims on preferences and purchase behavior – a systematic review. *Appetite*, *127*, 303–323.

Whalen, R., Harrold, J., Child, S., Halford, J., & Boyland, E. (2018). The health halo trend in uk television food advertising viewed by children: the rise of implicit and explicit health messaging in the promotion of unhealthy foods. *International journal of environmental research and public health*, *15*(3), 560–568.

Yarar, N., & Orth, U. R. (2018). Consumer lay theories on healthy nutrition: Aq methodology application in germany. *Appetite*, *120*, 145–157.