# Understanding individual food healthiness perceptions using a computational approach

by

**Natasha Gandhi**

**Thesis**

Submitted to the University of Warwick

for the degree of

**Doctor of Philosophy**

**WMG**

February 2022

THE UNIVERSITY OF WARWICK

# Contents

# List of Tables

# List of Figures

# Acknowledgments

*"I'm here to get it right, not to be right."*

- Brené Brown, *Atlas of the Heart*

First of all, I'd like to thank the teachers, lecturers, and mentors I've had throughout my life. I will always hold a special place in my heart for all those who have shown me kindness, shared their wisdom, and provided opportunities when it was above and beyond their job description. Thank you in particular to Dr Terry Dovey, who with the words "How would you feel about moving to Warwick?", changed my life in a phone call.

My biggest thanks goes to my PhD supervisors, Dr Lukasz Walasek and Prof Caroline Meyer. Thank you both for believing in me, encouraging me to aim high, normalizing journal rejection, setbacks, and all the uncomfortable feelings that come with the inevitable non-linear progress of a PhD. Caroline, thank you in particular for your enthusiasm for my research ideas, reassurance, and encouragement to always look at the bigger picture. Lukasz, thank you for all the guidance, time, and patience you have given me over these four years. You have taught me how to be a better writer, coder, and scientist, for which I cannot express enough gratitude. I could not have asked for better supervisors and role models, and feel so lucky to have been mentored by you both.

To my collaborators, Dr Sudeep Bhatia, Wanling Zou, and Piotr Bogdanski.

I cannot thank you all enough for everything you have taught me (especially with the understanding of the methodology and complicated statistics!), and for all the time you have spent helping make these papers as good as they can be.

To my examiners, peer-reviewers, annual reviewers, journal editors, and anyone who read and gave me constructive feedback on any part of my work. My aim has been to do this fascinating but complex research question justice, and I thank you for pointing out ways to make my writing or structure clearer, more relatable, and accessible to those who may end up reading it.

To my funders, EPSRC, this research would not have been possible without your financial support as well as your adaptability to the impact of COVID-19. Along similar lines, thank you to everyone at WMG who offered support, answered any questions I had, and gave regular updates.

Thank you to everyone who helped me problem-solve my many R or Python coding issues, whether they knew it or not. From colleagues, to those in the Stack Overflow community, R-ladies Coventry, and the many people who have put up free, and dare I say fun, resources online (especially Julia Silge and David Robinson!)

Thank you to everyone in the Behaviour and Wellbeing Science group at WMG (both past, present, and adopted into the team). I have been honored to learn so much from you all about your own areas of research and expertise, spanning multiple disciplines. I am grateful for your authenticity, willingness to help, and for making me feel as though I had a safe space to ask even the silliest of questions.

To my dance and fitness instructors, yoga and wellbeing teachers, and all those in Kenilworth that made it a home away from home, and who I never got to say goodbye to. Thank you for bringing your classes online so that I had some familiarity and routine to what felt like months of Groundhog Day. You always put a smile on my face, left me with moments of clarity, and allowed me to see small signs of progress somewhere, which motivated me in my working hours.

Thank you to my friends and family, of whom there are too many to name. Thank you for always checking in with me, being a phone call or WhatsApp message away when I've needed a distraction, or escape from my PhD bubble. If this pandemic has taught me anything, it is that time with you all is more fundamental to my wellbeing, and work productivity, than I could have previously realized. I love you all.

In particular, I'd like to give a special thanks to my mama, Rakesh Modi, for proof-reading every important thing I've ever written, often without much time, but always without question.

Thank you also to my dad, Anil Gandhi, for all your sacrifices, unconditional love, and support. Thank you for giving me the opportunity to be in a position where I can explore my curiosity and study what fascinates me the most.

To my mum, Mala Gandhi, thank you for always encouraging me to look past my fears and try, for being a shoulder to lean on, someone to learn from and look up to, and basically for being my biggest supporter in everything I do. I am lucky to have you as my best friend.

Last but not least, thank you to Kieran Compton, my partner. Thank you for being by my side for every confused thought and small win of my PhD roller coaster experience, even when we couldn't see each other in person. Thank you for being my rock, having confidence in me even when I doubted myself, motivating me when I was languishing, and just for all your love and support.

# Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor in Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below. Parts of this thesis have been accepted or submitted for publication by the author:

**Chapter 2:** Gandhi, N., Meyer, C., Bogdanski, P., Walasek, L. (2022) Computational Analysis of Superfood Representations in News Media. *Manuscript submitted for publication*

**Chapter 3:** Gandhi, N*., Zou, W*., Meyer, C., Bhatia, S., & Walasek, L. (in press). Computational Methods for Predicting and Understanding Food Judgment. *Psychological Science.*

Gandhi, N*., Zou, W*., Meyer, C., Bhatia, S., & Walasek, L. (2020). Knowledge Representations in Health Judgments. In *CogSci. Working paper.*

**Shared first authorship*

Parts of Chapter 3 have also been presented at the Subjective Probability Utility and Decision Making Conference (2019), Food GRP Interdisciplinary Conference (2020), and Food Standards Agency's Food for Thought Seminar (2021).

## Author Contributions

<u>Chapter 2</u>

**Natasha Gandhi:** Conceptualization, Methodology, Formal Analysis, Investigation, Writing - Original Draft, Visualization. **Caroline Meyer:** Supervision. **Piotr Bogdanski:** Methodology, Formal Analysis, Writing - Original Draft, Visualization. **Lukasz Walasek:** Conceptualization, Methodology, Writing - Review & Editing, Supervision.

<u>Chapter 3</u>

**Natasha Gandhi:** Conceptualization, Methodology, Investigation, Formal analysis of Studies 1B, 1C, 2A, 2B and 2C using existing code from Study 1A, Writing - Original Draft, Visualization. **Wanling Zou:** Formal analysis (Study 1A and all additional analyses in the Supplementary Materials; Validation of analyses for all other studies in the main manuscript. **Caroline Meyer:** Supervision. **Sudeep Bhatia:** Formal analysis for Study 1A and all additional analyses in the Supplementary Materials. Methodology. Writing - Review & Editing. **Lukasz Walasek:** Conceptualization, Methodology. Writing - Review & Editing. Supervision.

# Summary

Is granola healthy? What about steak? What knowledge do we use when judging the healthiness of different foods? Previous research recognizes that various non-nutrient related attributes are integral to food healthiness perceptions. However, our understanding of what contributes to healthiness judgments is constrained by researcher or participant assumptions. In this thesis, we uncover these attributes through a study of how different food names co-occur with other words in large-scale language data. Inspired by previous work using language data to predict judgments and choices, we use this data-driven methodology to reveal the psychological underpinnings of healthiness perceptions across all food categories.

Chapter 2 investigates the representation of superfoods in online news articles. By comparing articles written about the same 25 foods in a superfood context versus not, we were able to identify words uniquely used to characterize a food as a superfood. Our findings show that superfoods have the strongest association with health, far outweighing associations between organic and health, and even organic and naturalness. Moreover, mentions of the medicinal properties of these foods frequently occurred in a superfood context. Overall, the findings from this paper illustrate the complex, and often biased, representation of superfoods.

Chapter 3 demonstrates the consistently high ability of knowledge representations (taken from computational models) in predicting people's healthiness judgments of foods. This is found to be the case even if people are shown front-of-pack nutrient and calorie information. We also show how this approach can be used to uncover words and concepts that are more associated with healthy and unhealthy foods. Here, our results show that people's judgments of food healthiness are largely explained by the strength of association with naturalness and rawness. Conversely, we find environmental and social contextual factors are strongly associated with the model's predictions of unhealthiness.

# Chapter 1

# Introduction

## 1.1 Prologue

*Imagine you pick up your phone and check the news. While scrolling, you see a headline entitled "A rasher of bacon a day increases risk of cancer". This grabs your attention, so you open the news article. While reading, you come across words such as "latest study", "red meat", "harmful", "sausages", among many others. Health is a topic of interest to you, and so you regularly read articles about foods and their relation to health. For instance, on another day you read an article called "The lifesaving food 90% are not eating enough of", with references to "live longer", "fiber", "pulses such as beans, lentils and chickpeas". Now, you feel hungry, so you head to the shop for a sandwich. When deciding what to have, you narrow it down to your two favorite options: the "Chicken & Maple Cured Bacon Sandwich" and the "Falafel & Hummus Wrap". You go for the wrap, as it sounds like the healthier option. When making that healthiness judgment, however, you did not compare nutrient content information. If you had, you would have seen that the nutrient content for all key nutrients, except fiber, was worse for the wrap (based on nutrient composition values for real food products). To an outsider, they would likely conclude you either valued fiber content over the other nutrients, or health was not your primary motivator. However, instead, you used learned associations with the food names and product packaging to make this choice. Recent advances in computational science have raised the possibility of learning more about these associations through analyzing linguistic co-occurrences in large-scale natural language texts, like news articles. Thus, the aim of this thesis is to investigate the capability of such a* *computational approach in approximating what people universally associate with the intangible concept of food healthiness. Ultimately, this research will contribute to our understanding of naturalistic decision-making as well as food healthiness, with consequences for designing and evaluating related behavioral interventions.*

## 1.2 Relevance of Food Healthiness Perception Research

How can we help people make better, healthier, food choices? This overarching question dominates the psychology and behavioral science literature in the domain of food choice (Downs et al., 2009; Mazzocchi et al., 2015; Vecchio and Cavallo, 2019; Wyse et al., 2021). Consumers reveal remarkably accurate recollection of official generalized dietary advice (Laguna-Camacho and Booth, 2015; Paquette, 2005; Spiro and Wood, 2021), as well as a growing interest and preference for healthy eating (Plasek et al., 2020; Ronteltap et al., 2012). However, paradoxically, countries of all income levels are now facing a double burden of obesity and undernutrition (World Health Organisation, 2021).

Given the importance of food healthiness perceptions in making healthy food choices (Steinhauser and Hamm, 2018), there has been a substantial amount of research dedicated to understanding healthiness judgment errors (Carels et al., 2006; Gomez, 2013; Zou and Bhatia, 2021), as well as trying to correct inaccurate perceptions of commonly consumed foods (Borgmeier and Westenhoefer, 2009; Hagmann and Siegrist, 2020; Manippa et al., 2020; Vanderlee et al., 2021). Still, despite best efforts, consistently effective and evidenced-based approaches for helping consumers make well-informed and healthier food choices are yet to be found (Bandy et al., 2021; Brown et al., 2011; Gibson-Moore and Spiro, 2021).

## 1.3 The Complexity of Objective Food Healthiness

A caveat, of course, is that a food cannot be healthy or unhealthy (Rayner, 2017; Scarborough and Rayner, 2014). In fact, describing foods as "healthy" or in terms of "healthiness" has been contested in favor of "healthful(ness)", since foods themselves do not develop chronic diseases (Scarborough and Rayner, 2014). Nonetheless, these aforementioned terms will be used interchangeably in this thesis, as is often the case in the relevant literature and mainstream discourse.

Irrespective, a major challenge is that there is no gold standard measure for the healthfulness of single foods (Pinho-Gomes et al., 2021). Even among experts, there is much contention over how to establish food healthfulness objectively (Lobstein and Davies, 2009; Motoki et al., 2021; Nicklas et al., 2014; Plasek et al., 2020). For some, food healthfulness is synonymous with nutrient density (see Nicklas et al., 2014), others argue environmental and social aspects of well-being should also be considered (Oljans et al., 2018), and the majority emphasize that the healthfulness of individual foods should always be evaluated within the context of a person's total diet (Freeland-Graves and Nitzke, 2013; Plasek et al., 2020; Rayner, 2017; Scarborough and Rayner, 2014). That said, certified nutritional experts do agree that no single food or food group can provide us with all the nutrients we need for good health. Therefore, most experts are reluctant to rank the healthfulness of individual foods (Gaspar et al., 2020; Nicklas et al., 2014).

Instead, experts encourage certain nutrients and food groups to be eaten more (i.e., protein, fiber, fruit and vegetables, and nuts) and others to be eaten less (i.e., foods high in energy, saturated fats, sodium, and sugars). As such, the official recommendation across all countries for a healthy diet is balance, variety, and everything in moderation (de Ridder et al., 2017). This advice is encompassed within broad dietary guidelines, such as the EatWell Guide, to ensure recommended nutrient intakes are met (Freeland-Graves and Nitzke, 2013), while recognizing that a variety of different diets can meet nutritional needs (Green, 2015).

## 1.4 The Shift to Individual Food Healthiness

In recent years, however, there has been mounting pressure to classify or rank the healthfulness of individual foods (Pinho-Gomes et al., 2021). For one, consumers have communicated confusion over how to put broad dietary guidelines into practice (Lobstein and Davies, 2009). This is because most food choices involve decisions between foods within the same food category or product line, and some food items also contain multiple ingredients e.g. ready meals (Lobstein and Davies, 2009). Second, there has been a push to regulate marketing and advertising of foods high in fat, salt and sugar (HFSS foods), which requires a universal and objective healthiness criteria (Pinho-Gomes et al., 2021).

Therefore, for practical purposes, measures of individual food healthiness have been developed that reduce the concept of food healthiness to a food's nutritional composition and energy density (Lobstein and Davies, 2009; Nicklas et al., 2014; Santos et al., 2021; Scarborough et al., 2007a). Specifically, in this nutrient density approach to healthiness, these policy tools emphasize nutrients with the most evidence for increasing or decreasing the risk of non-communicable diseases such as Type-2 diabetes, high blood pressure, and some cancers (Jones et al., 2019).

## 1.5 A Nutrient-Focused Approach

### 1.5.1 Nutrient Labeling

Nutritional labeling is one of the most prominent policy tools worldwide for helping consumers make more healthful food choices (European Food Information Council, 2016; Kanter et al., 2018; van der Bend and Lissner, 2019). The premise is that visibility of nutrient content on food packaging will allow consumers to identify food items that are more nutrient-rich and lower in energy density (An et al., 2021). By informing consumers of nutrient content, and directing their attention towards this information, it is hoped they will be guided towards choosing the healthier option (An et al., 2021; Downs et al., 2009; Song et al., 2021).

At a minimum, back-of-pack nutrient labeling on food packaging must state energy (kCal and kJ), fat, saturates, sugar, salt, carbohydrates, and protein content (Ogundijo et al., 2021; Santos et al., 2021). Moreover, several supplementary nutrients (e.g., monounsaturated fats, polyunsaturated fats, and fiber) can also be voluntarily declared (EU Regulation No. 1169/2011).

However, due to the limited success of such back-of-pack nutrient labeling (Muller and Ruffieux, 2020; Rønnow, 2020), 31 countries have also implemented front-of-pack (FoP) nutrient labeling in line with World Health Organization (WHO) recommendations (Song et al., 2021). The purpose of all FoP nutrient labeling is to complement, summarize and simplify back-of-pack labeling information by highlighting energy (kCal and kJ), fat, saturated fat, sugar, and salt; the key nutrients from a population nutrition perspective (Jones et al., 2019; Santos et al., 2021).

### 1.5.2 Nutrient Profiling

In contrast, nutrient profiling models are a widely used policy tool that classifies or ranks foods according to their nutritional composition (Bandy et al., 2021; Scarborough et al., 2007a). Their specific purpose is to regulate both health and nutrient claims on food products, and food advertising aimed at children (Pinho-Gomes et al., 2021). Furthermore, nutrient profile models are used to decide how food items can be reformulated to increase their healthfulness, as well as which foods should be subject to health-related taxes (Rayner, 2017).

There are a variety of different nutrient density algorithms in use across different countries (Kissock et al., 2021; Nicklas et al., 2014). This is largely due to differing views on the standard units or serving sizes, but also what (and how many) nutrients should be included (see Nicklas, 2014). Nevertheless, if we take the UK-Ofcom nutrient profile model as an example, we can observe that the accepted key nutrients are similar to front-of-pack nutritional labeling (energy density, saturated fat, sugar, sodium), with fat content excluded but protein, fruit/vegetable/nut (as a percentage), and fiber content also included (Rayner, 2017).

## 1.6 Measures of Food Healthiness Perception Accuracy

Currently, most of our insights into lay people's healthiness perceptions are in reference to key nutrient content (Bucher et al., 2016; González-Vallejo et al., 2016; Mötteli et al., 2016; Rizk and Treat, 2014). As increasing adherence to recommended nutrient intakes is the goal, understanding healthiness perceptions of individual foods has also been studied using a corresponding nutrient-focused approach. Thus, judgment accuracy is typically assessed against scores calculated from nutrient profiling models (Bucher et al., 2016, 2015), or from the key nutrient criteria used in nutritional labeling and dietary guidelines (De Vlieger et al., 2017; Laguna-Camacho et al., 2018; Mötteli et al., 2016; Rizk and Treat, 2014).

Judgment discrepancies with objective measures have been commonly attributed to either lack of nutritional knowledge (Mötteli et al., 2016; Rizk and Treat, 2014), not using nutrient labels (Bucher et al., 2016), as well as overvaluing certain salient nutrients or keywords (e.g. "fruit") presumably taken from the food packaging (Bucher et al., 2015; Rizk and Treat, 2014; Sütterlin and Siegrist, 2015). However, these studies do not offer generalizable insights into healthiness perceptions as they are often focused on products from one food category (mainly cereals, snacks, or beverages) (Bucher et al., 2016; González-Vallejo et al., 2016), or one meal (typically breakfast) (De Vlieger et al., 2017; Laguna-Camacho et al., 2018; Sütterlin and Siegrist, 2015).

Evidently, the perceived importance of the different key nutrients varies depending on the food stimuli used, with existing regression-based studies highlighting any of fat, fiber, sugar, nut or seed, and fruit or vegetable content as significant predictors (Bucher et al., 2016; De Vlieger et al., 2017; Rizk and Treat, 2014). Nevertheless, these nutrient-based regression models (using various combinations of key contents) could only predict between 58%-71% of variance in healthiness judgments (Bucher et al., 2016, 2015). It should also be noted that these models were fitted to the entire dataset, without the use of cross-validation or separate testing data (de Rooij and Weeda, 2020). Taken together with the finding that there is no correlation between nutritional knowledge and reliance on certain key nutrients (Orquin, 2014; Rizk and Treat, 2014), it raises the question of whether other (non-nutrient) factors play a bigger role in food healthiness perceptions and judgments than currently assumed.

## 1.7 Limitations of Objective Food Healthiness Measures

Interestingly, experts state that nutrient profiling models are not a policy tool intended as dietary advice for consumers (Rayner, 2017; Scarborough and Rayner, 2014). This is because the presence of non-key nutrients and contextual information (e.g. foods consumed alongside a given food) should be important to healthiness perception accuracy (Scarborough and Rayner, 2014).

Indeed, Bucher et al. (2015) does acknowledge that some judgment discrepancies may be due to limitations of the nutrient profiling model as an objective measure. As an example, for smoked salmon, the researchers state consumers likely also consider omega-3 content, whereas the UK-Ofcom nutrient profile model does not take such supplementary nutrients into account. Similarly, in the same study, it was postulated that consumers may have overestimated the unhealthiness of McDonald's fries because of associations with the brand name, and other common accompanying foods like added salt and sauces. The UK-Ofcom nutrient profile model, on the other hand, categorized these fries as “healthy”.

In another study, participants were asked to serve themselves food for an entire day using a fake food buffet, containing 128 food replicas of perishable foods alongside 51 single-packet portions of real foods (Mötteli et al., 2016). Participants were randomly assigned to one of two conditions - to either demonstrate what they would normally eat (control group), or what they would choose for a healthy and balanced diet (healthy group). The results showed that consumers still exceeded recommended intakes of sugar and salt when they were purposefully making healthy food choices. As a result, Mötteli et al. (2016) recommended increasing nutritional knowledge in the public. However, upon closer inspection, it can be observed that the healthy choice group did choose fewer sweets, salty snacks, soft drinks, and fast food products than the control group. Instead, the high negative nutrient content was a result of consuming more foods like fruit juices, cheese, smoked salmon, and vegetarian products such as nuts, falafel, and salad dressing. Principally, these foods are from, or associated with, the main food groups provided in the broad dietary guidelines (Rayner, 2017). Thus, arguably, the aforementioned foods could make the total diet more healthy as intended (Nicklas et al., 2014; Rayner, 2017; Scarborough and Rayner, 2014).

Drawing conclusions about whether these foods are objectively healthier choices (not to mention what appropriate portion sizes might be) is outside the scope of this thesis, and part of a longstanding dilemma (see Nicklas et al., 2014; Scarborough & Rayner et al., 2014). Nevertheless, this apparent complexity demonstrates how the effectiveness of our current evaluation methods, and our understanding of consumer healthiness perception accuracy, are both not clear-cut.

## 1.8 Importance of Non-Nutrient Related Attributes

Even dietitians mention a relevance of attributes outside of key nutrients when judging the healthiness of individual foods (Scarborough et al., 2007b; Thurecht et al., 2018, 2020). For instance, the presence of “whole-grains” was the most important cue of healthiness for bread and breakfast cereals, whereas it was an absence of “non-nutritive additives” when making healthiness judgments for flavored yogurt (Thurecht et al., 2018). Indeed, dietitians also consider nutrients that are not necessarily provided on food packaging (e.g. “calcium”) as well as keywords like “fruit or vegetable”, “takeaway”, “fried” (Scarborough et al., 2007b).

Of particular interest is the study by Perkovic et al. (2021), who used a psychometric paradigm to explore food healthiness perceptions. This is a method that allows for quantitative representations of both nutrient and non-nutrient attributes. Recognizing that there are a wide range of attributes that may influence perceptions of food healthiness, Perkovic et al. (2021) presented participants with 43 common food items (from a range of food categories) and asked them to rate the healthiness of each of these foods using a seven-point Likert scale. Following this, the same participants were presented with the 43 food stimuli again, and asked to rate each of them on a predetermined set of characteristics (e.g., sugar content, fiber content, level of processing, origin (local vs nonlocal). In total, there were 17 characteristics, taken from the findings of a prestudy where (different) participants were asked what they considered relevant for judging food healthiness in an open-answer question format. To identify the most relevant attributes, Perkovic et al. (2021) used principal components analysis (PCA) to extract the reduced set of characteristics that best explain the common variance across the aggregate food healthiness ratings. In this study, perceptions of adolescents, adults and dietitians were compared, and considerable agreement was found in how they differentiated the healthiness of the food stimuli. For all three participant groups, perceived naturalness emerged as the strongest predictor of healthiness perceptions. This importance of naturalness, and related concepts such as “degree of processing”, “freshness”, “presence of food additives”, being “locally produced” or “homemade”, has also been found repeatedly in qualitative studies (Croll et al., 2001; De Vlieger et al., 2017; Paquette, 2005).

Nonetheless, it is worth highlighting that when a language analysis tool (Leximancer) was applied to expert and lay definitions of the term “nutritious” as part of a content analysis, there were 33 unique attributes that were identified as relevant to participant perceptions. (Bucher et al., 2017). Moreover, another survey-based study listed only 13 nutrient and non-nutrient based attributes when asking participants what they consider when making healthiness judgments (Lusk, 2019). As such, there appears to be a lack of consensus over what, and how many, attributes might contribute to food healthiness judgment formation, with some potentially still unknown.

## 1.9 Applying Multi-Attribute Choice Models to the Food Domain

Now consider the approximate 15,000 food and beverage products that can currently be found on Tesco's website (https://tesco.com/groceries), a leading supermarket in the UK. How might the average person make healthfulness judgments of all potential purchases? Even if one chooses to prioritize other motivators of food choice (sensory appeal, price, mood, convenience, familiarity, natural content, weight control, ethical concern), a healthfulness judgment of each item will still be made (Steptoe et al., 1995). In fact, there is much debate over the relationship between healthiness and tastiness perceptions, with references to an "unhealthy=tasty intuition" (Haasova and Florack, 2019; Raghunathan et al., 2006; Werle et al., 2013). Hence it is accepted that consumers demonstrate an ability to make healthfulness judgments of single foods regularly, rapidly, and with relative ease (Bui et al., 2017).

As a result, there is a subset of judgment and decision-making literature dedicated to understanding how food judgments and choices are made (Bhatia and Stewart, 2018; Scheibehenne et al., 2007; Schulte-Mecklenbeck et al., 2013). At the crux of this research is an investigation into the possible decision rules an individual might use to integrate or trade-off the many relevant, external attributes and cues available (Gigerenzer and Gaissmaier, 2011; Tversky and Kahneman, 1974; Weber and Johnson, 2009). The following two subsections summarize the two main theories of multi-attribute judgment and choice.

### 1.9.1 Weighted Additive Decision Rule

If the consumer is a rational decision-maker, they are expected to use the weighted additive rule to arrive at the most accurate judgment (Payne et al., 1993). This complex and effortful strategy requires the consumer to identify all relevant pieces of information (Shah & Oppenheimer, 2008).

Ideally, and as often assumed in existing studies (Ares et al., 2014; González-Vallejo et al., 2016; Van Der Merwe et al., 2010), this should be the government-provided nutrient content information. The consumer is expected to store, organize and recall the corresponding numerical values (key nutritional quantities) for each key nutrient, weighing up these nutrients according to subjective importance while integrating information for the alternative product options (Ettlin et al., 2015; Shah and Oppenheimer, 2008). It is at this stage that a healthiness judgment is made (Shah and Oppenheimer, 2008). The assumed use of this strategy underpins many of the commonly used front-of-pack nutrient labeling strategies from the monochrome reference intake design to traffic-light labeling (Grunert and Wills, 2007). In fact, support for evaluating nutritional information in this manner is provided by studies using self-report measures (Cowburn and Stockley, 2005). Slightly tangentially, consumers are also expected to retain and aggregate the reference intake information (provided as a percentage) of the four key nutrients in every food they consume per day (Lobstein et al., 2007; Yates, 2006). This is to ensure daily negative nutrient targets are not exceeded (Yates, 2006).

Nonetheless, front-of-pack nutrient labeling information is meant to be a benchmark, rather than the ground truth, as daily values vary based on each individual's caloric needs (Rayner et al., 2004; Yates, 2006). For practical purposes, standardization is necessary, and therefore all nutrient content guides are based on 2,000 calories, the recommended amount for the average woman (Rayner et al., 2004). Therefore, arguably, a rational decision-maker would also be expected to adjust the standardized reference intake percentages according to their caloric goals (Campos et al., 2011; Lobstein et al., 2007).

Irrespective, in real-world settings, it is considered doubtful that consumers have the time, cognitive resources, and computational capacity to compute these steps for every potential food purchase (Sanjari et al., 2017; Schulte-Mecklenbeck et al., 2013).

### 1.9.2 Heuristics

The alternative theory is that consumers rely on heuristics which are simplifying strategies (Gigerenzer and Gaissmaier, 2011; Payne et al., 1993; Tversky and Kahneman, 1974). Essentially, heuristics enable people to make decisions using only a few pieces of the available information (Scheibehenne et al., 2007).

To provide an example of the use of heuristics in the food domain, one can look to a known consumer reliance on packaging features from nutrition or health claims such as "rich in Vitamin C" (Davidović et al., 2021; Plasek et al., 2021), and "no added sugar" (Al-Ani et al., 2016), non-nutrient related labels like "organic" and “fairtrade” (Schuldt et al., 2012; Schuldt and Schwarz, 2010), to even just the brand name (Motoki et al., 2021). These attributes have been shown to increase healthiness perceptions of a product, despite alternatives often having no substantial differences in nutrient profile (Breen et al., 2020; Fernan et al., 2018; Kaur et al., 2016). The heuristic (or rather the subsequent cognitive bias) in this instance is often described as a health halo effect (Andrews et al., 2000; Chandon and Wansink, 2007), a derivative of the halo effect (Asch, 1961; Thorndike, 1920). A health halo effect has occurred when one positive attribute is incorrectly taken as an overall indicator of a product’s healthfulness (Andrews et al., 2000; Chandon and Wansink, 2007).

While the use of simple rules can allow people to make smart choices quickly (Todd and Gigerenzer, 2000), many food companies are taking advantage of the consumer desire for simplicity, and positioning their products as healthy by highlighting how they may have fewer “bad” ingredients (e.g. salt, sugar or fat) or more “good” ingredients (e.g. protein or fiber) than the norm (Huang and Lu, 2016).

Yet, these aforementioned marketing claims on food packaging do not increase healthfulness perceptions, or preferences, for every single food. (Steinhauser and Hamm, 2018). Rather, the familiarity of a given claim with that specific food item or food category is necessary to influence consumer behavior (Orquin and Scholderer, 2015; Steinhauser and Hamm, 2018). Indeed, various eye-tracking studies reveal that attentiveness to nutrition information, brand, and design elements of food packaging differs between food categories (Mhurchu et al., 2018; Orquin, 2014). In particular, nutrient labels, for example, were more likely to be viewed for convenience foods, cereals, snack foods, bread, and oils (Mhurchu et al., 2018); foods that are more ambiguous in healthiness. Instead, for the majority of foods, people rely mostly on their existing knowledge and expectations of the food’s healthiness, which is elicited through the food name or image on packaging (Schulte-Mecklenbeck et al., 2013).

This is a distinct difference from the majority of judgment and decision-making tasks, where choice sets are less familiar to the participant (e.g. choosing between electrical devices, houses, or jobs), with few expectations of the information relating to each attribute, or much awareness of other relevant attributes (Bhatia and Stewart, 2018). Indeed, researchers know that people apply this knowledge in the decision-making process of naturalistic choices, but seldom explore what this knowledge might be (Bhatia, 2017; Bhatia and Stewart, 2018). As such, we believe it is imperative to gain insights into how foods are represented in the mind to fully characterize perceptions and judgments of food healthiness.

## 1.10 Advances in Understanding Naturalistic Decision-Making

Faced with the conundrum that everyday objects and concepts are complex and require rich mental representations, judgment and decision-making researchers recently put forward a methodological approach to uncover these representations in a data-driven manner (Bhatia, 2017, 2019; Bhatia and Stewart, 2018). This has been achieved by connecting existing theories of judgment and decision-making with insights from other subfields of psychology and computational linguistics (Bhatia and Stewart, 2018). Essentially, it is accepted that heuristic-based judgments rely on the fundamental psychological process of association (Kahneman, 2002; Sloman, 1996). These associative judgment processes are understood to use inferences drawn from co-occurrence based statistical regularities from memory representations (Bhatia, 2017; Evans, 2008; Sloman, 1996; Tversky and Kahneman, 1974). However, until recently, no formal model existed that could quantify this process and draw predictions for naturalistic human behaviors (Bhatia, 2017).

Independently, associative processing had also been of interest to researchers studying semantic memory and language (Bullinaria and Levy, 2007; Jones et al., 2015). In these domains, there is a long history of using word co-occurrence statistics in natural language to successfully uncover word meaning, as well as semantic relationships (associations) between words (Firth, 1957; Harris, 1954). Motivated by this research, judgment and decision-making researchers realized you can considerably narrow down the number of unique, observable attributes associated with thousands of naturalistic choice objects by shifting the focus onto the systematic relationships that occur between these attributes, known as latent attributes. (Bhatia and Stewart, 2018). Simply put, the idea is that you could represent separate attributes such as "natural production", "presence of artificial additives", and "level of processing" using a single underlying theme, or latent attribute representation, of "naturalness".

As latent attributes are easier to retrieve from memory and incorporate, this strategy would minimize the cognitive effort required, but still allow for large numbers of attributes to be considered (Bhatia and Stewart, 2018). Indeed, Bhatia and Stewart (2018) found that using the weighted additive rule with latent attribute representations, rather than observable attributes, significantly outperformed the tested heuristic strategies for naturalistic multi-attribute food choices.

Typically, traditional methods of uncovering latent attributes require explicit participant responses such as rating the similarity of object pairs (e.g. multidimensional scaling), or the self-reporting of perceived important features (feature norms) (Richie and Bhatia, 2020). The structure of co-variability across similarity ratings reveal latent dimensions, which can be intuitively mapped to corresponding latent attribute representations (Bhatia, 2019). As you may recall, the study that currently offers the most insights into food healthiness perceptions made use of one of these traditional techniques (psychometric paradigm) in their approach (Perkovic et al., 2021). Nonetheless, these existing approaches only possess a few dimensions, and thus a limited number of latent attributes (Bhatia, 2019). In addition, the findings from traditional approaches are not generalizable to foods for which survey data does not exist. Of course, there are hundreds of attributes for hundreds of commonly consumed foods, with new or reformulated food products appearing regularly (Bhatia and Stewart, 2018). Therefore, it would be unrealistic (in terms of time and money) to attempt to obtain direct comparisons between all possible food pairs, or even feature norms for all commonly consumed foods (Bhatia, 2019; Bhatia et al., 2021a). Moreover, hundreds of latent attributes are needed to best mimic human judgments for any everyday object, which would require extremely rich representations (Bhatia and Stewart, 2018). Thus, traditional methods would be unable to scale to the number of latent attributes needed for accurate representations and judgment predictions (Richie and Bhatia, 2020).

Fortunately, relatively recent advances in machine learning and data science may offer a solution for approximating rich knowledge representations of foods (Mikolov et al., 2017, 2013; Pennington et al., 2014). While using the same premise as before, investigating co-occurrence based statistical regularities in natural language, the internet is instead used as the source of data (Aka and Bhatia, 2021; Bhatia, 2019; Bhatia et al., 2021a).

The underlying theoretical assumptions and computational techniques are known as *the distributional semantic hypothesis* and *vector space models* respectively. Importantly, however, we must emphasize that the purpose of vector space models is not to explain cognitive processes, but rather to act as a proxy for the associations stored in the minds of decision-makers (referred to as knowledge representations) (Bhatia et al., 2021a). In fact, predictive models would need to be trained on a much larger corpora size (billions of words) than most adults have been exposed to (Richie and Bhatia, 2020). If successful, there would no longer be a need for participant responses to uncover latent attribute representations for foods as richer insights could be generated from analysis of natural language texts (Richie et al., 2019).

Indeed, the recent accessibility of large quantities of natural language, with expansive vocabularies, has already led to highly accurate methods in other domains involving naturalistic behavior and high-level judgments (Bhatia and Bhatia, 2021; Bhatia, 2019; Richie et al., 2019). Nonetheless, the application of this approach as a means of understanding food healthiness representations is yet to be explored.

## 1.11 Outline of Thesis

In this thesis, we focus on approximating food representations from online news articles, as news sources most likely reflect the collective beliefs and representations of everyday objects and concepts (Bhatia et al., 2021b). More importantly, while references to nutritional content are likely to be captured in news articles, equally likely are associations with other similar foods and non-nutrient words. For instance, “fatty” and “salty” may frequently appear in articles about “bacon” but so might words like “processed”, “sausages”, “pigs”, and “fry-up”. Hence, this computational text analysis approach has the potential to encompass non-quantifiable food attributes, triangulating the existing qualitative and quantitative research on food healthiness. Thus, our goal is to establish the capability of numerous theoretically similar, but computationally different, approaches in explaining healthiness perceptions of individual foods. In both empirical chapters, we do not have any hypotheses as our goal is to maximize the predictive accuracy of the computational models.

In Chapter 2, our aim is to explore representations of a sub-group of foods (superfoods) that are popularized in mainstream media discourse due to their high association with healthiness. All of the computational techniques used in this chapter apply a “bag-of-words” approach to uncovering representations. Essentially, “bag-of-words” models reveal the most frequent co-occurrences of words in a group of articles, but disregard word order and other grammatical features (Gefen et al., 2017). As we will demonstrate, while word meaning is not captured in its entirety, this approach can be very informative and offers more intuitive results than those using richer representations as obtained in Chapter 3.

In Chapter 3, our intention is to create a formal, predictive, model that can provide quantitative healthiness judgment predictions for a wide range of foods. To capture the rich and nuanced representations of foods as required for this task, we use a popular subclass of vector space models known as word embedding models. We use word embedding models, specifically word2vec as our main model, because they provide us with extremely high-dimensional vectors. These word embedding models have successfully predicted low-level psychological phenomenon such as human similarity judgments, free association, and categorization (Günther et al., 2019; Jones et al., 2015; Lenci, 2018). More recently, researchers have used the representations from these word embedding models as predictors in regression models for human judgments (Bhatia, 2019; Bhatia et al., 2021a; Richie et al., 2019). This requires participant ratings in order to train the model and then the predictive ability is tested on out-of-sample ratings, using cross-validation (Aka and Bhatia, 2021; Bhatia, 2019; Bhatia et al., 2021a). Using this approach, multiple studies reveal that the rich representations obtained from vector space models successfully predict participant ratings for multiple concepts, even though participants made judgments when only given names of stimuli (Bhatia, 2019; Bhatia et al., 2021a). Consequently, we use off-the-shelf (pre-trained) vector space models, in the same manner as the existing research in other domains (Richie and Bhatia, 2020). As such, we will not specify the technical details and assumptions for training these models in this thesis (instead see Mikolov et al., 2013; Mikolov et al., 2017; Pennington et al., 2014; Pereira et al., 2016). Our main goal is to replicate the methodology from these high-level judgment studies to this previously unexplored domain, but also extend the research by adding a series of robustness checks (using expert judgments, food images, and testing model generalizability). This is in addition to including comparative predictive models that use nutritional composition data (Nutrient Model, Combined Model and two extended versions of both in the Supplementary Materials). Furthermore, we will also offer a theoretical contribution by exploring how judgments made with quantitative numerical data alongside the food name (taken from nutritional labeling strategies in the UK) might shift the predictive ability of each of our models.

Next, in Chapter 4, I summarize the research findings of the previous two chapters. I also discuss the theoretical and practical implications of this research, with recommendations for future research provided.

Finally, in Chapter 5, I give a reflective account of the additional analyses and considerations that were not included in my main chapters.

### 1.11.1 Chronological Order

The structure of the chapters in this thesis does not reflect the order in which they were written. Chronologically, Chapter 3 was written first and, after a series of revisions, was published online in *Psychological Science* on March 17, 2022. The findings of Chapter 3 informed the direction of Chapter 2 and the focus on exploring superfood representations. At present, this paper has not been published. Notes for the "Introduction" and "Discussion" chapters were written throughout, with the write-up for both being predominately written during the final six months of the thesis period. Lastly, the final "Further Reflections" chapter was written in response to feedback from the viva examiners.

### 1.11.2 Data and Code Availability

Please note that de-identified (and permitted) data and code to replicate the findings from all chapters in this thesis have been made publicly available via GitHub. It can be accessed at https://github.com/ngandhi95/Food-Healthiness-Perceptions-Thesis-Code.git

# Chapter 2

# Superfood Representation in News Media

## 2.1 Abstract

What does blueberry, avocado, quinoa, and ginger have in common? These food items are often regarded as superfoods because of the assumed benefits these items have for one's health and well-being. These perceptions persist despite the fact that many health benefits of superfoods are not supported by evidence. One route by which incorrect beliefs about superfoods may be perpetuated and maintained is through news media. In this chapter, we set out to investigate how foods claimed to be superfoods are represented in the discourse of online news. We use computational language models to extract the unique topics and terms used to discuss superfoods over a ten year period, and compare those to descriptions of superfoods provided by a sample of survey respondents. Our results show that news coverage is dominated by many specific claims about the healing properties of superfoods. Terms occurring in the context of such foods often refer to specific vitamins and minerals, and diseases (e.g. cancer, heart disease). Results of the structural topic model further demonstrate that articles mentioning superfoods are more likely to include topics about a) nutrients, physical appearance and health in the same context, b) retail strategies, and c) scientific research about the health benefits of superfoods. Overall, representations of superfoods uncovered by our analysis are well aligned with people's perceptions. Together, our results demonstrate a potential source of strong and misplaced association between superfoods and health.

## 2.2 Introduction

Despite the fact that many people strive to eat more healthily, there is little consensus about the *what, when, and how* of a healthy diet. Although plenty of guidance and recommendations from government and experts exists (Julia et al., 2021), discourse in online media is now dominated by diverse opinions and advice concerning what people "should" be eating to improve their health and well-being. Given that people's perceptions and behavior are shaped by the media representation of food healthiness (Nagler, 2014; Oakes and Slotterback, 2001), it is essential to better understand what messages about food healthiness online media perpetuate.

It is known that food marketing influences consumers' perceptions of food healthiness (Chandon and Wansink, 2012; Plasek et al., 2021; World Health Organization, 2021). In fact, food and beverage companies invest heavily in healthy food marketing in response to consumer demand (Samoggia et al., 2020). The most recognized and researched type of food marketing strategies are those found on food packaging, such as health claims and symbols, package design, and branding (Plasek et al., 2020; Silchenko et al., 2020). However, large food and drink companies are also positioning themselves as nutritional educators (Garcia and Proffitt, 2021), often marketing specific "healthy" product categories and ingredients (Chandon and Wansink, 2012; Mintel, 2016). As a result, the online media discourse surrounding foods also contains marketing concepts disguised as healthy eating advice (MacGregor et al., 2021; Samoggia et al., 2020).

These online media messages have led to the marketing term "superfood" being popularized in everyday discourse (Delicato et al., 2019; Roth and Zawadzki, 2018), which is the focus of this chapter. Broadly speaking, superfoods refer to foods that are naturally rich in macro-nutrients, as well as various vitamins and minerals (Jagdale et al., 2021). This is despite no clear evidence linking a given food item to any health outcomes (Cloutier et al., 2013; Siipi, 2013; Thurecht et al., 2018). Moreover, the superfood narrative in the media is at odds with the advice of nutritional experts, who advocate a view of healthiness in terms of overall dietary patterns (Freeland-Graves and Nitzke, 2013; Lobstein and Davies, 2009; Lusk, 2019). In 2007 the European Food Safety Authority banned the word "superfood" on food packaging unless accompanied by a specific permitted health claim (EU Regulation No. 1924/2006). Likewise in America and Australia, as all food claims must be supported by scientific evidence, superfood is not a claim that would be authorised for use on food packaging (Food Standards Australia New Zealand, 2018; US Food And Drug Administration, 2016). Yet despite this regulation, perceptions about the superiority of superfoods persist, which is likely to originate from the abundance of mentions in magazine and news articles, blogs, and on social media channels (Liu et al., 2021; MacGregor et al., 2021). Consequently, representations and implicit beliefs about certain foods having a superfood status are still being reinforced, leaving consumers misinformed (Delicato et al., 2019). Brands are then able to take advantage of this loophole by adding putative superfoods to the ingredient list of snack items (Breen et al., 2020), baked goods (Meyerding et al., 2018), and beverages (Brownbill et al., 2020), as a means of increasing healthiness perceptions and sales of their products (Meyerding et al., 2018; Mintel, 2016).

As superfood is a term often used to promote food items that are from the main food groups (e.g., fruit and vegetables) (Jagdale et al., 2021), one may wonder whether there is any harm to this practice. At a minimum, the superfood discourse feeds into an overly simplistic view of foods as "good" or "bad", causing people to abandon efforts to improve their overall diet (Freeland-Graves and Nitzke, 2013). Use of such language can also contribute to the prevalence of disordered eating (Douma et al., 2021), where individuals follow strict, restricted diets and experience guilt after perceived food transgressions in pursuit of healthiness (Galfano et al., 2022). Furthermore, the classification of various foods as superfoods can produce a health halo effect (Amos et al., 2019), whereby consumers incorrectly interpret the health benefits and health risks of certain foods (Breen et al., 2020). Indeed, one side effect is often for consumers to underestimate calorie consumption of healthy foods (Carels et al., 2006, 2007; Larkin and Martin, 2016; Provencher et al., 2009), resulting in weight gain and therefore an increased risk of diet-related illnesses.

From a psychological standpoint, little is known about what underpins people's representations of superfoods. Most academic research in this area uses case studies of known superfoods to investigate either consumer demand for superfood products (Graeff-Hönninger and Khajehei, 2019; Groeniger et al., 2017; Loyer, 2016; Meyerding et al., 2018), nutritional evidence for popular health and nutrient claims (Jagdale et al., 2021; Proestos, 2018; Šamec et al., 2019; Štepec et al., 2020), or the environmental and social consequences from the rise in superfood consumption (Bedoya-Perales et al., 2018; Magrach and Sanz, 2020; Reisman, 2020). Of the studies that have explored perceptions and beliefs surrounding superfoods, participants were provided with pre-defined survey measures (Franco Lucas et al., 2021; Liu et al., 2021; Rojas-Rivas et al., 2019), meaning that our understanding is potentially constrained by the researchers' expectations and hypotheses.

As a complementary approach to the existing literature, here we undertake an exploratory computational analysis of natural language corpora. Specifically, we analyze superfood representations in online news articles. To date, only two papers have looked at the coverage of superfoods in news articles (MacGregor et al., 2021; Weitkamp and Eidsvaag, 2014), but for different purposes. Weitkamp and Eidsvaag (2014) investigated the influence of scientists on the information reported in news articles about superfoods, and MacGregor et al. (2021) used critical discourse analysis to explore how the marketing of superfoods in the media promotes neoliberal ideologies.

A similar methodological approach has been used to investigate the relationship between various foods and perceived health benefits, however, the superfood status associated with these foods in the media discourse has been overlooked. Moreover, these studies were conducted using social media sources such as Twitter (Kāle and Agbozo, 2020; Lynn et al., 2020; Pilař et al., 2021; Samoggia et al., 2020; Vidal et al., 2015), or Reddit (Blackburn et al., 2018). For instance, one study investigating perceptions of kale using Twitter data between January 2020 and June 2020 found frequent mentions of health-related concepts (e.g. "anti-inflammatory", "immune boost") and a lack of reference to taste or hedonistic concepts in its research (Kāle and Agbozo, 2020). However, the question remains about whether this finding is reproducible when analysing the discourse in mainstream news media, as well as if it would generalize to other known superfoods.

In this chapter, we leverage the latest methods from natural language processing (NLP) to analyze online news articles about known superfoods in different contexts. Our chosen corpus of online news articles (News on the Web—NOW) captures the representation of superfoods in the US and UK media over a ten year time period (2010-2020). As such, this approach allows us access to the actual discourse being used to shape widespread public opinion on superfoods. Our study focuses on a sample of 25 food names, identified by a sample of survey respondents, resulting in 872 articles that mention at least one of these foods and the word "superfood" (superfood articles) and a further 56,981 articles about the same foods with no mention of the word "superfood" (non-superfood articles). Such a large number of relevant articles means that we can uncover meaningful patterns about superfood representation in the media. We conduct three analyzes on our large corpus of online news to evaluate the nature of discourse pertaining to superfoods. First, we explore commonalities in language between superfood articles and descriptions of superfoods provided by participants who completed our survey. Next, we use two computational techniques to establish predictive features of articles written about superfoods. The first of these computational techniques is text classification. For this analysis, we formally compare articles written about our 25 chosen foods in a superfood context with closely matched articles written in a non-superfood context. Therefore, our classifier is forced to rely on subtle differences in language when making predictions as to whether an article is about superfoods or not. An advantage of this approach is that we can uncover the most important words, used as predictors by the trained classifier, to distinguish the representation of a food in a superfood specific context. Last, in order to gain further insight into the interaction between various words and concepts related to superfoods in online news media, we used structural topic modeling. Topic modeling allows us to formally identify latent themes and topics in our sample of news articles about the 25 foods. We use structural topic modeling (STM) specifically as it allows us to compare the likelihood of each topic appearing in superfood or non-superfood articles. Finally, as an additional comparison, we replicated both of the above-mentioned computational analysis techniques using "organic" in place of "superfood" mentioning articles. The purpose of this is to ascertain whether we can successfully capture the unique representation of superfoods, rather than a general concept of healthiness. Ultimately, the goal of our computational analyzes is to learn which food attributes, associations, and concepts are emphasized in the representation of a food as a superfood in online news media.

## 2.3 Methods

### 2.3.1 Corpus Selection

The corpus of online news articles used in this chapter were taken from the NOW Corpus (http://corpus.byu.edu/now/), which is a collection of online newspaper and magazine articles, maintained by Mark Davies at Brigham Young University. This is the only English-speaking corpus that is larger than a billion words (Davies, 2017), and so was most suitable for exploring a relatively niche topic like superfoods. The meta data for this corpus includes "article ID", "word count", "date", "country", "news outlet", "url" and "title". Specifically, for our analysis, we used a static local copy of the NOW Corpus, accessed in May 2020, which covers the time period between January 2010 to February 2020. Only articles published in the United

States of America and Great Britain were analyzed.

### 2.3.2 Identifying Food Names

We compiled a list of food names by downloading data from the U.S. Department of Agriculture (2019) Food Composition Data, and McCance and Widdowson's Composition of Foods Integrated Dataset (Public Health England, 2021), as both are the official sources of information about commonly consumed foods in the USA and UK respectively. This list was used in an initial preprocessing step to filter articles only containing food names.

### 2.3.3 Participant Survey Data

We conducted a short online survey on Prolific Academic to obtain a list of most recognized superfoods. Details of this study are provided in the Supplementary Materials, as here we focus only on the use of the 25 most commonly identified superfoods that we used to select online news articles for our computational analyzes.

### 2.3.4 Data Preprocessing and Article Selection

All steps to extract and clean the relevant sample of online news articles were performed using R (version 4.1.0) (R Core Team, 2021). The R package "spacyr" (Benoit and Matsuo, 2020), was used for data preprocessing. We started with a sample size of 13,871,016 news articles published in the United States of America and Great Britain during the time period specified. First, we removed HTML tags, URL links, non-alphabetical characters (e.g. special characters, numbers, and all punctuation except for hyphens), and standardized all text and titles of the news articles to lowercase. Then we selected only news articles that contained the word "food" or "diet" either in the text or title. At this stage there were 779,919 news articles. We then removed all articles that either had duplicated article IDs or that had identical article titles from the same news outlet. Following this, we filtered articles that mentioned at least one of the food names in our food name list (see Identifying Food Names section for detail), resulting in 547,568 food-related news articles. Our next data preprocessing steps involved tokenization (splitting article texts into single word units), removing stop words (frequent words that provide little information e.g. "and", "in" and "the"), and part of speech tagging (to identify a word's function within a sentence, e.g., "green" as an adjective). Part of speech tagging allowed us to select the most meaningful parts of speech (nouns, adjectives, and verbs) from the articles. We also performed lemmatization, turning words representing the same concepts into its base form (e.g., "energizing" and "energized" both became "energize"). Next, we replaced all possible spellings of the word superfood (e.g. "superfood", "super-food", "super foods") with the former spelling for consistency. This then allowed us to identify the 1,169 articles that mentioned the word "superfood" at least once in the article text or title, and the 216,769 that did not. Consequently, we further subset the data to only include articles that specifically mentioned the 25 superfood names given by participants in the online survey mentioned above. This resulted in a final sample of 57,853 news articles, with 872 articles that mentioned the word superfood at least once (henceforth referred to as superfood articles). As a last step, all articles were then tagged with a dummy variable to denote their status as either a superfood or non-superfood article.

### 2.3.5 Computational Methods

#### Word Frequency: A Bag-of-Words Model

As an initial exploration of our corpus, we used a word counting technique known as a bag-of-words model to provide a simplified representation of our superfood articles sample (see Kowsari et al., 2019). The bag-of-words approach was chosen over other known frequency statistics (such as tf_idf and weighted logs odd ratio) because these alternative methods would give priority to unique and obscure words even when found in only a small number of articles. Instead, we were interested in finding the most common words across all articles in our chosen context.

For this analysis, we included additional data preprocessing steps. The same steps were replicated when preprocessing participants' survey responses to two questions asking them to describe or define a superfood, for comparative purposes. Further details are provided in the Supplementary Materials.

### Text Classification

Text classification is a supervised machine learning technique that allows us to predict whether an article is written in a superfood context or not. An offshoot of this approach is that we can subsequently identify the terms that underlie the classifier's predictions. As a consequence, text classification can enable us to make inferences about concepts that are most likely to be associated with superfoods in a sample of articles discussing the same group of foods.

For text classification, it is optimal to have a balanced distribution of articles in each group. However, our dataset was highly unbalanced, with the superfood articles making up only 1% of the sample. In addition, the much larger 'non-superfood' class likely included articles that only scarcely referenced the food items of interest. Therefore to down-sample the non-superfood articles and balance the corpora, we used propensity score matching (Ho et al., 2011; Rubin and Rosenbaum, 1985). We obtained the propensity scores for each sample using a generalized linear model with a logit link function. We regressed the article class (i.e., superfood mentioning or not) on the counts of the top 25 superfoods selected by survey participants. The premise was that the logit model could estimate the probability of each article being a superfood or non-superfood article, and the predicted probabilities (propensity scores) would reveal how likely each non-superfood article could serve as a viable counterfactual (or replacement) for that superfood article. All superfood articles were paired with it's nearest neighbor; the non-superfood article with the closest propensity score. Note that each match was independent of each other, and thus the same non-superfood article could be matched to several superfood articles (greedy matching). All unmatched non-superfood articles were then discarded from the sample, allowing for maximum homogeneity between the two comparison groups and a reduced sample size. In total there were 872 pairs of articles in our text classification sample, 10% of which were always held-out for model evaluation during cross-validation (as explained below).

Next, we trained a text classification model to discriminate between the superfood and non-superfood articles in our balanced sample. Specifically, we used the logistic regression classifier (also known as the maximum entropy classifier); a linear model often used for text classification tasks due to its interpretability. We chose to encode the data with unigrams, bigrams, and trigrams. Each type of an n-gram is an expression of N consecutive words combined. Previous studies have shown that increasing the n-gram range leads to better performance in variety of text classification tasks (Bharadwaj and Shao, 2019; Shah et al., 2018). We represented the data using a document-feature matrix, where each column represented an n-gram, each row represented a news article, and the observations for each of the documents corresponded with the count of the word's occurrence. As an example, the single hypothetical sentence "Superfoods are very good", would correspond to N columns in the document-feature matrix: "superfoods", "are", "very", "good", "superfoods are", "are very", "very good", "superfoods are very", "are very good". Also note that L1 penalty was imposed on the norm of the coefficients, which resulted in majority of irrelevant coefficients shrinking towards 0, to avoid over-fitting.[1]

When training the regularized logistic regression, it was important to choose the optimal strength of the penalty imposed on the L1 norm of the model's coefficients. We tuned the regularization strength parameter by running a grid search over a range of values and evaluating the average F1 score on 10-fold cross-validation split for each of them. The F1 score is the harmonic average of the model's precision and recall, further discussed in the results section. The entire fitting procedure was implemented in the "glmnet" package in R (Simon et al., 2011).

As an additional analysis, we repeated all the aforementioned steps replacing "superfood" with the word "organic". The reason was to ascertain whether the representation of superfoods is unique or reflects an overarching perception of food healthiness. Specifically, using the sample of 25 superfoods given by participants, we marked news articles that mentioned the word "organic" vs. those that did not.

### Topic Modeling

Finally, in order to examine whether the differences identified by the classifier generalized to more broad, latent themes found in the entire subset of the news articles corpus we used the Structural Topic Model (STM) (Roberts et al., 2016, 2014). Topic modeling refers to a family of unsupervised statistical learning techniques that identify underlying latent semantic structures characterized by frequent occurrence of a vocabulary subset in corpora of natural texts. In the past, topic modeling has been applied to analyze corpora from variety of areas, such as social media dis-

[1] Formally, logistic regression solves the following optimization problem: $\min_{(\beta_0,\beta)\in R^{p+1}} -\left[\frac{1}{N}\sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T\beta) - \log(1+e^{(\beta_0+x_i^T\beta)})\right]$, while the L1-regularized variant changes that to $\min_{(\beta_0,\beta)\in R^{p+1}} -\left[\frac{1}{N}\sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T\beta) - \log(1+e^{(\beta_0+x_i^T\beta)})\right] + \lambda|\beta\|_1$, where $\lambda\|\beta\|_1$ is the regularization penalty imposed on the model's coefficients.

course (Zamani et al., 2020), financial news (Bybee et al., 2020) or historical texts (Barron et al., 2018).

Topic modeling relies on several assumptions, which enabled us to extract topics from our newspaper corpus. One such assumption is that each document is composed of a mixture of topics, and each topic is formed using a probability distribution of words. In the same manner as the bag of words model, it also assumes that there is no order to the words in a document, and that documents are independent. The distinguishing feature of STM is that, while building on the basic idea of probabilistic topic modeling, it allowed us to incorporate document-level covariates (or metadata) into the model's structure (Roberts et al., 2019). STM was therefore most appropriate for our goal of quantifying the effect of our dummy variable (superfood article or not) on the topical structure of the news articles. As such, STM allowed us to identify topics present across all articles that were more prominent when the term "superfood" was used.

In order to facilitate model estimation, we further subset our data to only include articles mentioning any of the superfood items specified by the survey participants more than twice, resulting in a sample of 18,219 non-superfood articles and 577 superfood articles. Since the number of the latent topics to be estimated has to be specified a priori, we conduct a grid-search over a range of values and chose the highest number (K = 12) that offered a notable improvement in the exclusivity score calculated over the top 10 words in each of the topics. A word is said to be exclusive with respect to a given topic if it has a high probability of appearing in the topic and low probability of appearing in the other topics estimated by the model (Roberts et al., 2014). Exclusivity of a model is the aggregated exclusivity of the top N words for each of the topics. The details of the parameter search can be found in Figure 2.5 of the Supplementary Materials.

## 2.4 Results

### 2.4.1 Superfood Names

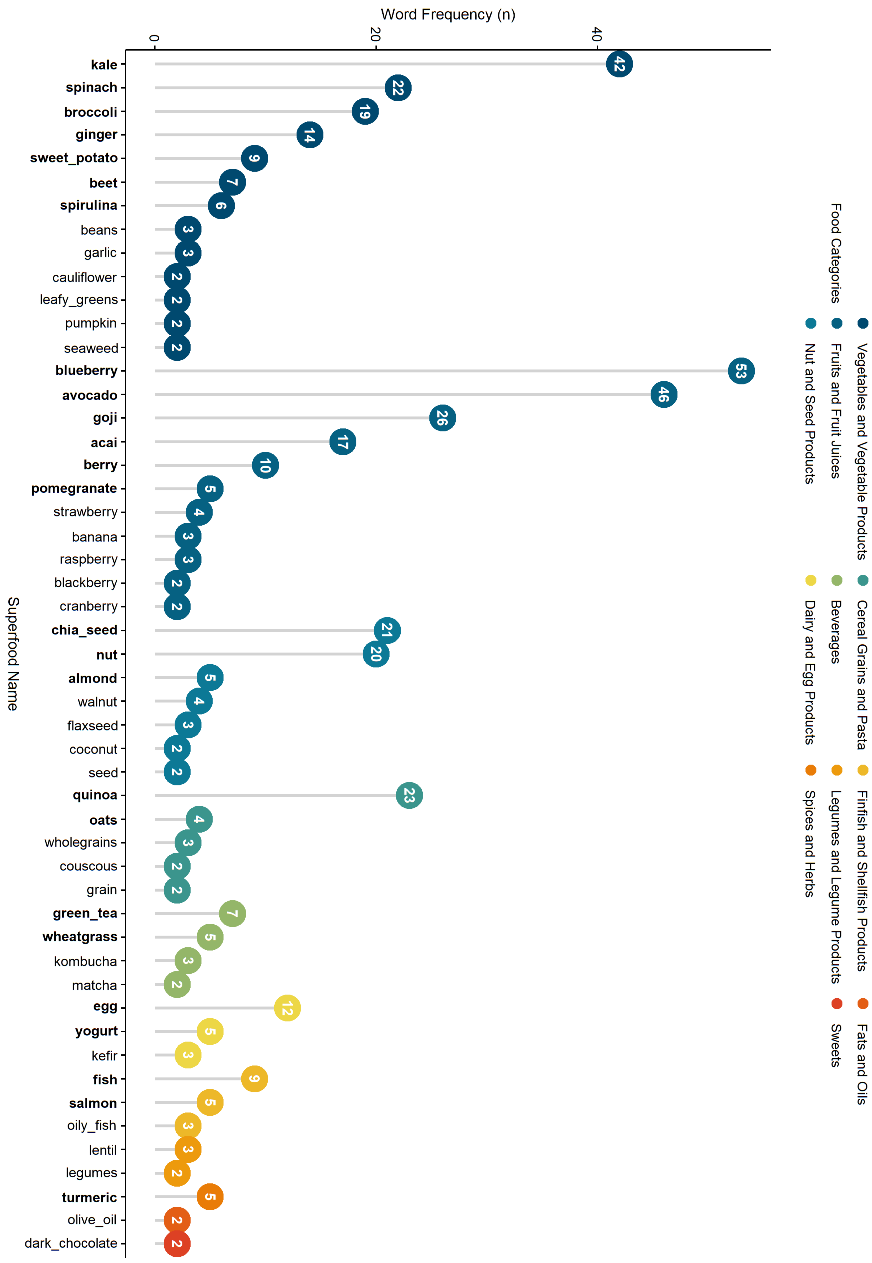



**Figure 2.1** All food items, that participants associate with the "superfood" label, mentioned more than once (using word frequency).

In total, there were 115 unique superfood names listed by our survey participants (see the OSF repository associated with this project for the full list: https://tinyurl.com/SuperfoodStudy), with 51 given by at least two participants. As can be seen from Figure 2.1, these 51 foods named as superfoods belong to a wide range of food categories from Vegetables and Vegetable Products (e.g., kale and spinach) to Spices and Herbs (e.g., ginger and turmeric), and Sweets (e.g., dark chocolate). Unsurprisingly, the majority of these foods were from the "Vegetables and Vegetable Products" category and "Fruit and Fruit Juices" category with "blueberry", "avocado", and "kale" being the most frequently mentioned (53, 46 and 42 mentions, respectively). The top 25 food names mentioned by participants, and the food names subsequently used for the computational analysis, are highlighted in bold on the X-axis of Figure 2.1. These foods are listed in descending order of word frequency and include blueberry, avocado, kale, goji, quinoa, spinach, chia seed, nut, broccoli, acai, ginger, egg, berry, fish, sweet potato, beet, green tea, spirulina, almond, pomegranate, salmon, turmeric, wheatgrass, yogurt, and oats.

### 2.4.2 Word Frequency

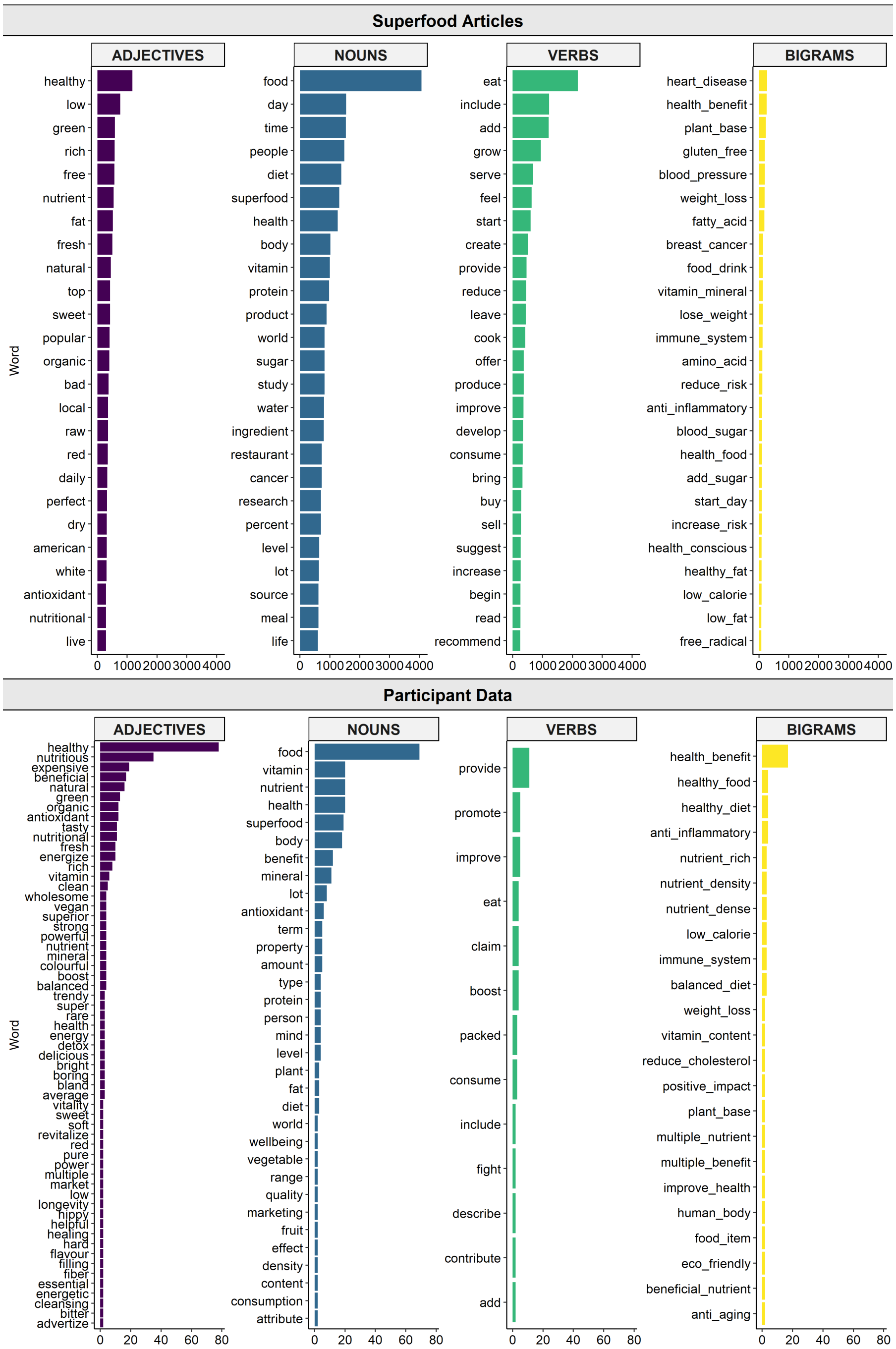



**Figure 2.2** A comparison of the top 25 adjectives, nouns, verbs and bigrams taken from online news articles written in a superfood context with the adjectives, nouns, verbs, and bigrams mentioned by more than one participant to describe or define a superfood.

Our first analysis concerns the distribution of most frequent words in superfood articles. These are presented in Figure 2.2 alongside data from our survey participants, separately for adjectives, nouns, verbs, and bigrams. Even at a glance, the concept of health is most prominent, but we can also see several references to the sensory properties of foods, naturalness, weight control, and scientific research.

The relationship between superfoods and health is demonstrated by the disproportionate prevalence of health related terms in superfood articles. Explicit references to health related terms (e.g. "healthy" (N = 1,179), "health" (N = 1,261), and "health_benefit" (N = 261)) are evident in the adjectives, nouns, and bigrams plots generated from the superfood article data. Broader connections to health were also made with mentions of positive and negative nutrients such as "vitamin" (N = 1,001), "protein" (N = 976), "sugar" (N = 822), "antioxidant(s)" (N = 296), "fatty_acid(s)" (N = 184), "amino_acid(s)" (N = 115), and "healthy_fat" (N = 89). Relatedly, a link with specific diseases is apparent in the media coverage, with an abundance of mentions of "cancer" (N = 725), "breast_cancer" (N = 142), and "heart_disease" (N = 278). Moreover, texts included numerous references to medical concepts and terms such as "blood_pressure" (N = 195), "immune_system" (N = 124), "anti_inflammatory" (N = 107), and "blood_sugar" (N = 105). Together with common bigrams such as "reduce_risk" (N = 113), "increase_risk" (N = 96), and verbs (e.g. "provide" (N = 465), "reduce" (N = 447), "offer" (N = 372), "improve" (N = 359) and "increase" (N = 265)), these words highlight a discourse where superfoods are likened to medicine.

Reference to the sensory properties of superfoods can be seen from the most frequently used adjectives in superfood articles. Color is most mentioned, with "green" (N = 592) the third most frequent adjective, and red (N = 349) and white (N = 314) also present. Surprisingly, the only taste descriptor in the top 25 adjectives was "sweet" (N = 426). Instead the majority of adjectives in superfood articles refer to the concept of naturalness (e.g. "fresh" (N = 508), "natural" (N = 453), "organic" (N = 406), "raw" (N = 361)). The theme of weight loss is most apparent when looking at bigrams, with frequent collocations in the superfood articles including "weight_loss" (N = 194), "lose_weight" (N = 127), and "low_calories" (N = 85). Moreover, we can also see words ("study" (N = 820), "research" (N = 703), "percent" (N = 701)), alluding to the use of scientific evidence as a persuasive technique for the promotion of superfoods in the media.

When comparing these frequent words from the superfood articles to the participant data, we find a similar prevalence of concepts. First, the healthiness aspect of superfoods is most emphasized, with "healthy" (N = 78) being the top adjective, many references to health benefits and medically related terms (e.g., "anti_inflammatory" (N = 4), "immune_system" (N =3) and "reduce_cholesterol" (n = 2), and multiple nutrients named (e.g., "vitamin", "antioxidant", "protein"). Similarly to the superfood articles, the most common terms for sensory properties reflected the appearance of superfoods ("green" (N = 13), "red" (N = 2), "bright" (N = 2) and the more general "colorful" (N = 2)). Taste was also referred to, but both in a positive and negative way (e.g., "tasty" (N = 11), "delicious" (N = 3), "bland" (N = 3)). Participants also associate superfoods with naturalness, with 16 individuals mentioning the term "natural" explicitly. Plus, other naturalness related words such as "organic" (N = 12), "fresh" (N = 10), "clean" (N = 5), "wholesome" (N = 4) and "pure" (N = 2), were mentioned. A few participants also referred to aspects of weight control. For example, "weight_loss" (N = 2) was explicitly mentioned but also "low_calorie" content (N = 3), "detox" (N = 3) and "cleansing" (N = 2), which are often used in the context of weight loss. Interestingly, no science or research terms were referenced more than once, suggesting that participants may either not remember this persuasion technique or deem it important.

To uncover the specific words that underlie predictions about an article being about superfoods, we used text classification. Note that the use of propensity score matching means each of the superfood articles was compared against a non-superfood article that was highly similar in context (determined by similar counts of the top 25 foods selected by survey participants). For robustness purposes, we replicated this analysis comparing articles that mentioned the word "organic" (organic articles) vs those that did not (non-organic articles). Table 2.1 shows the performance of the classifier model on the out-of-sample dataset. We summarize the results in terms of three commonly used classification metrics - accuracy, precision and recall. Accuracy is simply the number of cases predicted to be in their true respective class. Precision refers to the number of examples correctly predicted as belonging to a given class, as a proportion of all examples belonging to that respective class. Recall (also known as sensitivity) represents the number of observations correctly predicted as a given class divided by the total number of observations actually belonging to that class. As seen in the Table, our classifier has an accuracy rate of 67% for superfood articles and 72% for organic articles, meaning that it is able to classify these online news articles better than chance. As a result, we can conclude that our classifier model can sufficiently pick up linguistic differences between articles written about foods in a superfood context or not, as well as in an organic context or not.

### 2.4.3 Text Classification

**Table 2.1** Text classification validation performance

| | Precision | Recall | Train size | Test size | Accuracy |
|---|---|---|---|---|---|
| Non-superfoods | 0.663 | 0.701 | 785 | 87 | 0.672 |
| Superfoods | 0.683 | 0.644 | 785 | 87 | |
| Non-organic | 0.688 | 0.803 | 4156 | 462 | 0.720 |
| Organic | 0.764 | 0.636 | 4156 | 462 | |

A benefit of our approach is that we can use the classifier model to find the n-grams with the highest probability of being in a news article classified as a superfood article or an organic food article. Figure 2.3 presents the top 100 n-grams, scaled to be proportional to the log-odds of the corresponding coefficients for each group.

Looking first at the superfood classifier n-grams in Figure 2.3, the first thing to note is an overlap between the n-grams produced using this approach and the words obtained from the simpler word-counting technique in Figure 2.2. The relationship between superfoods and health is yet again prevalent. For example, many n-grams reference the nutritional composition of foods such as "source (of) potassium", "high protein", "nutrient rich", "mineral (and) vitamin", "vitamin c (and) e", "monounsaturated fat", "low sugar", "anti-oxidant", and "good/bad cholesterol". The classifier also picks up on the discourse surrounding superfoods and illnesses, with n-grams predictive of superfood articles including "heart disease percent", "cancer percent/fight", "boost immune system", "carcinogen" and "reduce stress".

Additionally the unigram "premature" is highly predictive but it includes two uses of the term in two separate contexts: premature death (for 16 of the superfood articles) and premature ageing (for 13 of the superfood articles). Other n-grams that show a relationship with beauty include "protect skin", "facial", "hydrating", "regime", all of which were not seen in the word frequency (bag-of-words) analysis.

In terms of sensory attributes, "tasty" is the only n-gram predictive of a superfood article, and the only reference to natural content is "unpasteurised" or "grass feed". We also detect nuances in language such as the paradoxical nature of foods marketed as superfoods, with predictive n-grams indicating a local origin ("locally sourced") but also their exotic nature ("ancient", "Peruvian" and "South America").

In comparison, there is little crossover between n-grams predictive of superfood articles and organic foods, despite the fact that both terms are typically perceived as cues to healthiness. Instead, articles written in the context of organic foods are predominately centered around food production methods, the concept of naturalness, and foods' environmental impact. Nonetheless, the most predictive n-gram was "certified", suggesting that articles about organic foods may highlight the necessary government standards and regulations that must be met for a food to be labelled organic. This is followed by "natural food", which along with "wholesome", "pure", "whole food" as well as "decay" and "rot", suggest a strong link between organic foods and naturalness. Similarly, there is an emphasis on the local aspect of foods with n-grams like "locally source(d)", "local food/farm", "small farm", and "farmer'(s) market" also being predictive of news articles about organic foods. Convenience or ease of access is also implied, as "delivery service", "open restaurant" and "metropolitan" may suggest. Environmental and ethical impact is another theme that stands out, with direct references to "eco/environmentally friendly", "sustainable food", "ethical", "good quality" and "cooperative". Moreover, while health benefits of foods are alluded to with descriptions like "healthful" and "medicinal", the focus is on the presence or absence of chemicals (e.g. "fertiliser", "pollutant", "contaminant"). From this, we can conclude that our model successfully picks up on representations of healthiness that are unique to superfoods, rather than all health-related terms like organic.

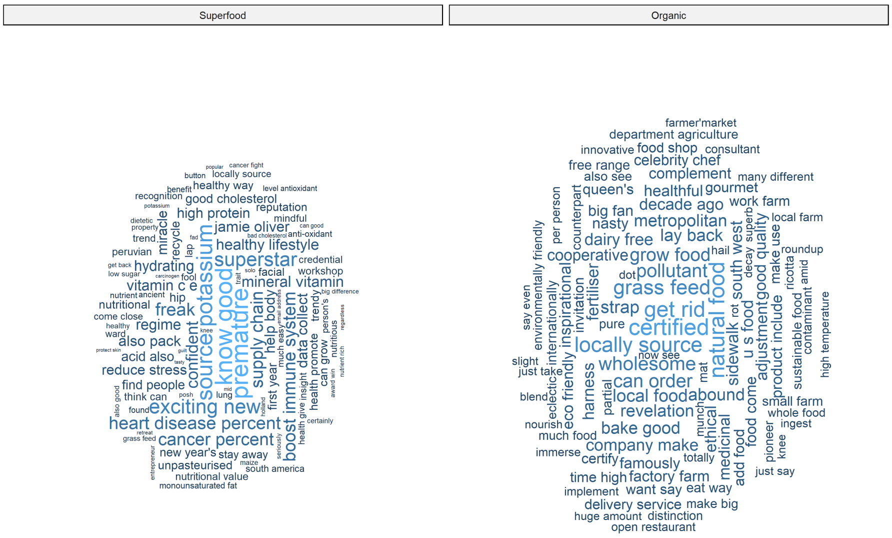



**Figure 2.3** Wordcloud of the 100 n-grams most predictive of the terms "superfood" and "organic" occurring in the sample of online news articles.

### 2.4.4 Topic Modeling

To confirm and extend the findings from the text classification, we used STM to extract topics. Essentially, STM enables an automated discovery of differences in latent themes and topics between articles written about the same 25 foods from different contexts. Figure 2.4 summarizes the estimated effect of the presence of the "superfood" or "organic" term on the difference in proportion of a latent topic appearing in the articles. A positive value on the X-axis indicates a larger prevalence of a given topic in superfood articles (Panel A) or organic articles (Panel B). Our two reported topic models were fit with the same 12 topics as determined by the grid search (see Methods for detail). Topic labels were assigned by using the top 10 keywords most associated with the given topic.

In articles written about the same 25 foods, what group of words (constituting a topic) is most likely to occur when the term "superfood" is mentioned? As shown in Panel A of Figure 2.4, the topic relating to diet and weight clearly stands out, with a 25.72% (95% CI [23.20%, 28.15%]) higher likelihood of occurring in superfood articles. The words most indicative of this topic were "diet, fat, healthy, help, health, protein, vitamin, weight, body, and sugar". Notably, this topic includes both terms related to food nutrients ("vitamin", "protein", and arguably "fat") and appearance ("weight", "shape"). The fact that these terms appear alongside "health" as one of the most representative words, suggests that the model detected a relationship in discourse between diet, appearance and health. In comparison, Panel B shows that this topic was the third most prevalent in organic articles relative to non-organic articles. Moreover, it was only 2.90% (95% CI [1.78%, 4.03%]) more likely to occur in an organic context. Given that this topic is considerably more prominent in the discussion around superfoods, one could infer that this association is pivotal in the representation of superfoods in the media.

Perhaps less surprising was the higher likelihood of retailing concepts appearing in superfood articles relative to non-superfood articles about the same foods. However, although this topic was the second most prevalent in superfood articles, the mean difference was much smaller compared with the first (diet, appearance and weight) topic at 0.02 (95% CI [0.00, 0.04]). The 10 words that formed our interpretation of this topic consisted of "product, company, market, store, sell, uk, business, price, buy, and consumer". This retailing topic was also the second most likely to appear in organic food articles in proportion to non-organic food articles, but was slightly more likely to occur in organic articles than superfood articles (mean difference of 0.04, 95% CI [0.03, 0.05]).

The third topic more likely (by 1.94%, 95% CI [0.04, 3.84%]) to appear in superfood articles vs non-superfood articles was one associated with scientific research. Keywords constituting this topic such as "study", "disease", "cancer", "health", and "research" had also been present in the findings of our previously mentioned computational analysis techniques. Again, having the word "health" within the top 10 words of this topic suggests a discourse where scientific evidence is given to suggest a relationship with these foods and health or disease. Interestingly, despite more debate and scientific research conducted to assess the relationship between organic foods and health, this topic pertaining to research was less likely (by -0.97%, 95% CI [-1.92%, 0.03%]) to occur in organic news articles than non-organic articles. Moreover, this contrast in discourse demonstrates another distinction between representations of superfoods and organic foods in the media.

Our topic modeling approach also reveals that the concept of cooking, "cook, add, minute, recipe, salt, heat, serve, water, bowl, pan", was slightly more likely to be found in articles written about superfoods, by 0.75% (95% CI [-1.14%, 2.71%]).

This was the topic with the largest difference in ranking between superfood and organic contexts, shown from a visual comparison between the superfood (Panel A) and organic (Panel B) plots. We can also see that this cooking topic was ranked higher than the topic pertaining to eating out "restaurant, menu, dish, chef, serve, wine, bar, open, taste and meal", a topic found to be less likely to appear in superfood articles than non-superfood articles (by -3.26%, 95% CI [-5.36%, -1.05%). Thus, these findings suggest superfoods are more likely to be promoted as ingredients for home cooking, rather than as a treat when dining out. Conversely, the opposite is true for organic foods, where organic is more likely (by 0.84%, 95% CI [-0.37%, 2.03%) to be discussed in the same context as eating out, but less likely (by -2.28%, 95% CI [-3.38%, -1.23%]) to be mentioned in references to recipes.

Contrary to expectations, a topic relating to naturalness was only 0.34% (95% CI [-1.03%, 1.90%]) more likely to occur in superfood articles relative to non-superfood contexts. The words used to define this topic were "farm, grow, plant, water, farmer, produce, crop, animal, feed, production". On the other hand, this topic was ranked first for likelihood (at 8.30%, 95% CI [7.41%, 9.17%]) in organic vs non-organic articles, supporting the assumption that this finding is due to representation differences from these two contexts. It is also worth noting that the keywords relating to naturalness all appear to relate specifically to food production methods, with terms relating to the rawness or purity not detected by the topic model.

One may also wonder about the topics least likely to occur about these 25 foods in a superfood context, and how this differs to topics least likely to occur in an organic context. Figure 2.4 shows us that the majority of the topics (7 out of the 12) were found in articles where the term superfood was not mentioned, with the same being true for the organic comparison. Of these topics, the one referring to the fishing industry was least likely (mean difference of -0.05, 95% CI [-0.07, -0.04]) to be mentioned in a superfood context, defined by the topic label "water, river, lake, sea, fishing, catch, boat, fisherman, whale, ocean". This was also true of the topic least likely to occur in articles about organic foods relative to non-organic foods, but with a mean difference of -0.03 (95% CI [-0.04%, -0.02%]). The second least likely topic in superfoods articles (mean difference of -0.05, 95% CI [-0.06, -0.04]) consisted of words such as "state, photo, family, city, school, community, country", which appears to reflect a discussion of social factors surrounding consumption of our sample of foods. The same topic was ranked relatively similarly (third least likely) in the organic comparison (mean difference of -0.02, 95% CI [-0.03, -0.02]). Interestingly, the third least likely topic (mean difference of -0.05, 95% CI [-0.06,

-0.04]) to occur in a superfood context demonstrates a focus on the unsustainability of fishing practices and environmental consequences (consisting of unique words such as “climate, change, fishery, population”). Along similar lines was a topic referencing human-wildlife coexistence (mean difference of -0.04, 95% CI [-0.06, -0.03]) with the words most representative of the topic being “animal, bird, specie(s), human, bear, female, live, wildlife, male, insect”. These two topics were also less likely to occur in an organic context but had a slightly higher mean difference (-0.02, 95% CI [-0.03, -0.01] and -0.02, 95% CI [-0.03, -0.01] respectively) in comparison to superfoods vs non-superfood articles. The next topic slightly more likely to be mentioned outside of a superfood context (by -4.02%, 95% CI [-5.82%, -2.32%]), seems to reflect a discourse relating to food tourism, with words including “island, hotel, local, city, place, beach, town, old, around, visit”. For organic articles, this food tourism topic was equally likely to occur relative to non-organic articles (mean difference of 0.00, 95% CI [-0.01, 0.01]). Lastly was a more abstract topic of words denoting the writer’s personal perspective (“think, want, really, life, back, feel, love, never”), which had a small mean difference of -0.04 (95% CI [-0.05%, -0.03%]). However, by comparison this same topic was the second least likely to appear in an organic context (mean difference of -0.03, 95% CI -0.04, -0.02]).

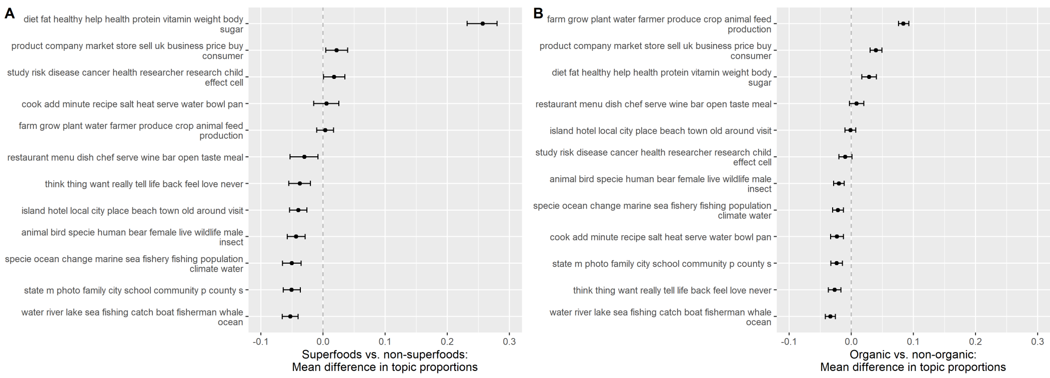



**Figure 2.4** Estimated differences (and 95% confidence interval) in topic probabilities between superfood and non-superfood articles (left panel - A), and organic food and non-organic food articles (right panel - B). Each plot is ordered by the topics most prevalent in the superfood or organic articles.

## 2.5 General Discussion

Our computational approach makes a number of contributions to our understanding of superfood representation in the online news media. Through a series of comparisons (with participant data, and with articles about the same foods in either a non-superfood context or organic context) we extracted the words, concepts, and topics most strongly associated with the term superfood. First, we found a unique emphasis on the relationship between individual foods and health benefits in superfood articles. Second, we observe a distinct use of medical terminology (such as “cancer”, “immune system”, “heart disease”, “risk”) in a superfood context, which is notably absent in the representation of organic foods. Third, against our own expectations, terms stressing the naturalness and environmental impact of these foods were infrequent in the characterization of superfoods. As a whole, considering superfood has no official definition, our findings offer a deeper understanding into the concept of superfoods. Given the role media plays in shaping people’s beliefs and attitudes, our results also contribute to our understanding of the origins of misconceptions concerning the health and well-being benefits of superfoods.

Although a link between superfoods and health benefits is expected, this finding provides support of previous research findings in a data driven manner (Franco Lucas et al., 2021; Loyer, 2016; Rojas-Rivas et al., 2019). All three of our bottom-up approaches found “health”, “healthy” and specific nutrients such as “protein”, “sugar”, “vitamin” to have the largest association with the term “superfood”. Such consistency of findings between our three bottom-up techniques demonstrates the robustness of a computational approach in uncovering superfood representations. Indeed, drawing attention to isolated compounds present in foods is not new to superfood advertising (Scrinis, 2013); in fact it was a highly successful marketing strategy on food packaging until unfounded nutrient claims were banned in the 1990s (Goldberg and Sliwa, 2011; Silchenko et al., 2020). Nonetheless, a strength of our topic modeling approach is that we can now see the extent of the association between this nutrient-focused conceptualization of health and superfoods. As such, we observe how the representation of superfood with health considerably exceeds the relationship of organic and health, and even organic and naturalness in media discourse. While we cannot use present results to claim that this superfood representation in the media directly influences consumer perceptions, similar language in participant responses does suggest individuals are at least aware of the same association. Furthermore, we found that mentions of health, various nutrients, weight, and appearance, emerge within the same topic (taken from our topic modeling analysis). This implies a discourse where superfoods are touted for weight loss as part of health messaging (Rodney, 2018; Sikka, 2019), despite scientific evidence establishing that weight is a poor indicator of health (Frederick et al., 2020; Saguy and Almeling, 2008). Considering nutrient-focused marketing detracts from the recommended "total diet approach" to healthy eating (Freeland-Graves and Nitzke, 2013), plus the implications of a suggested linear relationship between weight and health for disordered eating behaviors (Frederick et al., 2020; Pilař et al., 2021), our findings highlight a need to further extend health and nutrition claim regulation to the online media marketing of food items.

Our results also draw attention to a medicinal representation of foods, centered around disease prevention, which is more likely to occur in a superfood context. Cancer is the most frequently associated disease in our article sample about superfoods, with a slight emphasis on breast cancer. However, in the words of Cancer Research UK (2020) "there is no good evidence that any one food prevents cancer, including superfoods". Most of the research conducted on individual foods that are reported in the media are either from animal studies (Jagdale et al., 2021), in vitro (outside of a living organism) studies (Šamec et al., 2019), or from single studies that should be interpreted with caution (Ladher, 2016). This is also true for the other diseases mentioned in our superfood articles (e.g. heart disease). Thus, the ability of our untrained model to identify a representation of superfoods based on weak evidence (Inoue-Choi et al., 2013), reinforces the role of the media in creating confusion about healthy eating (Hackman and Moe, 1999; Nagler, 2014; Weitkamp and Eidsvaag, 2014). Again, as evident from the language reflected in participant responses, the superfood discourse in the media may therefore help explain the discrepancy between people's inaccurate beliefs about food healthiness and official dietary guidance.

Contrary to the entwined relationship between health and naturalness found in previous research (Loyer, 2016; Michel et al., 2021; Perkovic et al., 2021; Roman et al., 2017; Siipi, 2013), naturalness was not a concept stressed in the online news article representation of superfoods. This is more surprising because naturalness representation was detected in participant responses, as well as in the comparative analysis of organic food articles. One possible explanation is that the relationship between chosen superfoods and naturalness can be assumed by design, whereas the same food can be sold as organic or not and thus naturalness associations would need highlighting in organic discourse. Another factor to consider is that superfoods are often sold in supplement form, involving high levels of processing, and so claims about their naturalness may appear contradictory.

The general lack of coverage concerning the social and environmental consequences of superfoods in online news articles may explain why some participants perceived superfoods as “eco-friendly”. However, the existing scientific literature on superfoods reports a detrimental social and environmental impact due to increased demand worldwide (Bedoya-Perales et al., 2018; Loyer, 2016; Magrach and Sanz, 2020). It is perhaps unsurprising that superfoods are spun in a positive light within a media marketing discourse, even if this unbalanced representation further enhances the halo effect of superfoods. However, given the relatively broad range of news articles from a variety of news outlets in this study, one might expect a higher prevalence of this topic in superfood articles than revealed in our topic modeling analysis. As a result, it would be interesting to explore whether differences occur between different news outlets, and if this finding is also true of representations in social media. Moreover, as consumers demonstrate a preference for environmentally-friendly foods (Franco Lucas et al., 2021), a recommendation for future research is to assess how increased awareness of environmental consequences from global scale production of superfoods (e.g. water depletion, soil degradation, reduction in biodiversity, and carbon footprint) might influence perceptions, preferences, and purchase behavior for superfoods.

Our relatively small number of superfood articles, both initially (1,169) and after selecting only articles mentioning 25 known superfoods (872), is unlikely to capture the entirety of online superfood news articles written between January 2010 and February 2020. We chose to prioritize minimizing researcher influence, selecting articles from our corpus using arbitrary means (count of the word “superfood”) rather than adding further news articles from specific news outlets (e.g. The New York Times, or The Guardian). It is also worth noting the existence of related marketing terms that have spawned from the superfood discourse (e.g., “superfruit”, “supergrain”, and “super berry”) (Liu et al., 2021; Loyer, 2016). Unfortunately, too few articles were available in the NOW corpus to extract meaningful themes and representations. Nonetheless, despite some limitations, our approach captures meaningful patterns that are consistent with discourse findings about known superfoods using other corpora (Kāle and Agbozo, 2020)

Overall, we believe that the strength of our approach is that we can uncover and quantify the unique representations of superfoods in the news media. While the term superfood is banned on food packaging, here we demonstrate how this term is prevalent outside of the supermarket environment. More importantly, we demonstrate a number of unique dimensions that make up the representation of superfoods in the media. The next stage for researchers is to ascertain the extent to which these representations influence food perceptions, and ultimately food choices. For now, we recommend advertising regulatory bodies to pay close attention to the loopholes being used to produce these, misleading, and potentially harmful associations.

## 2.6 Supplementary Materials

### 2.6.1 Participant Survey

Here, we provide additional details about the participant online survey used to ascertain a list of known superfoods as well as participant definitions and descriptions for the term itself. Ethical approval for this study was obtained from the University of Warwick's Biomedical and Scientific Research Ethics Sub-Committee.

#### Participants

One hundred participants were recruited via Prolific Academic in return for a fixed payment of £0.63 (£7.77/hour). Only English-speaking adults from the United Kingdom or the United States of America were eligible to take part in our survey. Four participants who answered "no" when asked if they were familiar with the term superfood were removed from subsequent analysis, leaving 96 participants (69% female) in our sample. On average, participants were 35.40 years old (SD = 13.48; range = 18-77). Of these participants, 75% had no dietary restrictions.

#### Design and Procedure

Participants completed a short online survey, consisting of an eligibility question followed by three main questions and three demographic questions. All questions were presented in the same order and read: 1) "Are you familiar with the term superfood (yes or no)?", 2) "Name at least 5 superfoods", 3) "List at least 5 adjectives that you associate with superfoods" and, 4) "In your own words, how would you define a superfood". Participants were then asked about their age, gender and dietary restrictions (options provided were "Vegan", "Vegetarian", "Other (please specify if you wish)" and "None of the Above").

#### Data Analysis

The list of 25 superfoods (as well as non-food words) was determined using word frequency, identical to the computational method outlined in the section "Word Frequency: A Bag-of-Words Model".

### Data Preprocessing

The same data preprocessing steps as for the superfood articles (tokenization, stop word removal, part of speech tagging, lemmatization, and conversion to lowercase) were applied to the participant data, for analysis at the aggregate level. In addition, as part of the standardization for food names, all spaces between open compound words (e.g. peanut butter and sweet potato) were replaced with an underscore. Moreover, bigrams from participants' open text responses (to question 4) were identified as words that appeared next to each other in a sentence more than once. Furthermore, we chose to concatenate all text given by the same participant to account for, and remove, repeated words (e.g. if the same participant gave the word "healthy" as a response to both question 4 and 5 it was only counted once). Spelling mistakes and inconsistencies, either introduced by participants or through the lemmatization process (e.g. "brocolli" to "broccoli" and "slimme" to "slimming"), were identified using the "hunspell" package in R and corrected manually (Ooms, 2020). Spellings were manually standardized to US spelling to avoid duplication. Only words tagged as adjectives, nouns, verbs (and manually as bigrams) were kept for the analysis of non-food words.

## 2.6.2 Additional Data Preprocessing Steps of Superfood Articles

As mentioned in the main text, we included additional data preprocessing steps for the Word Frequency (bag-of-words model) analysis only. First, we created a manual list of stop words, allowing us to remove all food names from the 872 superfood articles (see the Identifying Food Names section in main text for detail). Second, we identified bigrams as words that appeared next to each other in a sentence within the corpora 50 times or more, and replaced spaces between these words with underscores. Next, we tokenized and lemmatized the article text, also removing stop words and tagging parts of speech. Similarly to the participant data, all spelling mistakes were identified using the "hunspell" package in R and corrected manually (Ooms, 2020). We then removed sparse terms, which are the words that occur in less than 3% of the 872 superfood articles. This allowed us to plot the most common terms of the whole superfood corpora, rather than those than rarely occur. Subsequently, we filtered the most meaningful parts of speech (adjectives, nouns, verbs and bigrams). Finally, we subset the top 25 words for each of these part of speech tags, which can be seen, alongside the comparative participant data, in

Figure 2.2 of the main text.

### 2.6.3 Topic Modeling: Grid Search

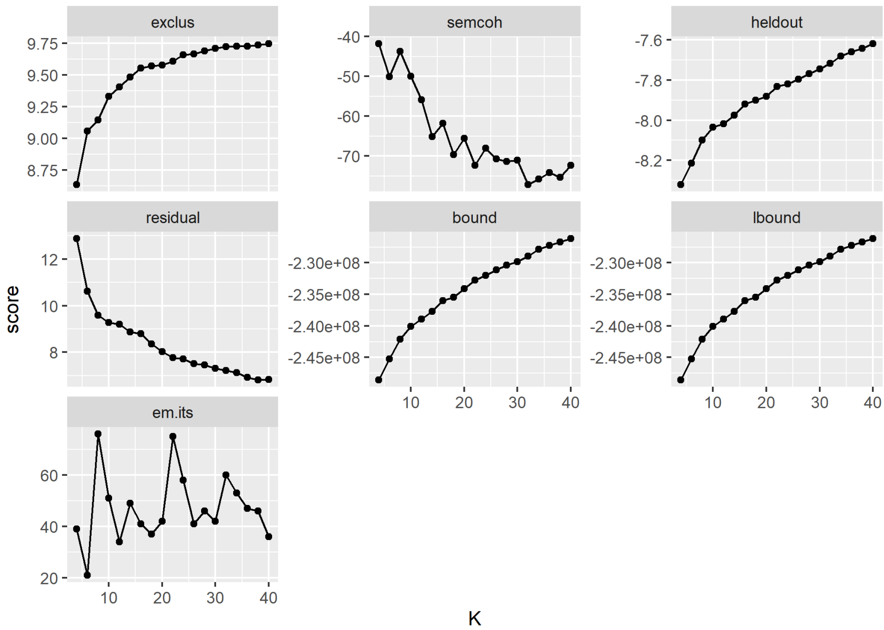



**Figure 2.5** Grid search evaluation results. The model was optimized based on the exclusivity score.

As referenced in the Topic Modeling section of the Computational Methods Section, Figure 2.5 shows the measures we used to find the optimal number of latent topics. We searched K-values between 4 and 40, shown on the X-axis of Figure 2.5. From left to right, the plots refer to exclusivity, semantic coherence, heldout likelihood, residual, bound, lower bound, and finally “em.its” refers to the total number of EM iterations used in fitting the model (Roberts et al., 2019). Here, our model was set to run at a maximum of 100 iterations. We optimized the model on exclusivity, which compares word distributions between topics to determine the likelihood of the top words of one topic being top words in the other topics (Roberts et al., 2019). Another important measure is semantic coherence, introduced by Mimno et al. (2011), which refers to the probability that the top words of one topic co-occur within our corpora of superfood articles (Pandur et al., 2020). We also consider held-out likelihood estimation, similar to cross-validation, which checks the model’s predictive performance by estimating the probability of words occurring within a document after they have been removed (Pandur et al., 2020). Measuring residuals is useful for determining how much variance remains at a given topic number, and whether more topics would be needed to account for any overdispersion. The bound is a measure of convergence, with the model considered converged when there is a small enough change between iterations (Pandur et al., 2020). The lower bound simply applies a correction to the bound so that the bounds are directly comparable (Roberts et al., 2019).

# Chapter 3

# Predicting Food Healthiness Judgments

## 3.1 Abstract

People make subjective judgments about the healthiness of different foods every day, which in turn influence their food choices and health outcomes. Despite their importance, there are few quantitative theories about the psychological underpinnings of such judgments. This study introduces a novel computational approach that can approximate people's knowledge representations for thousands of common foods. We use these representations to predict how both lay decision-makers (general population) and experts judge the healthiness of individual foods. We also apply our method to predict the impact of behavioral interventions such as the provision of front-of-pack nutrient and calorie information. Across multiple studies with data from 846 adults, our models achieve very high accuracy rates ($r^2$ from 0.65 to 0.77), and significantly outperform competing models based on factual nutritional content. These results illustrate how new computational methods applied to established psychological theory can be used to better predict, understand, and influence health behavior.

## 3.2 Introduction

Poor diet is one of the primary preventable causes of premature death in high-income countries (Bauer et al., 2014). Understandably, people want to consume healthy foods as they recognize the relationship between diet and health. However, people can only make healthy food choices to the extent that they can correctly judge a food's healthiness. One obstacle to healthy eating is that there is no normative answer to the question: "what makes food healthy or unhealthy?" (Lobstein and Davies, 2009). Still, it is commonly believed that food healthiness judgments are strongly linked to beliefs about the nutritional content of food products (Scarborough et al., 2007a).

Indeed, health organizations worldwide routinely emphasize which nutrients people should avoid (high saturates, fats, sugars, salt) and which they should consume more of (high protein, fiber) (Lobstein and Davies, 2009). This is apparent in the design of numerous front-of-pack food labeling formats, which attempt to simplify complex nutrient information for consumers. Such interventions highlight overall energy content and the presence of nutrients that are most associated with the rising rates of obesity and chronic diseases (Kanter et al., 2018). Yet, evidence about the effectiveness of such interventions is mixed (Sanjari et al., 2017).

The success of front-of-pack labeling rests on the assumption that people rely on energy and nutrient information to judge a food's healthiness (Orquin, 2014). However, evidence suggests that healthiness judgments reflect pre-existing knowledge that people associate with foods' perceived naturalness (Siipi, 2013) and taste (Turnwald et al., 2017). These are further influenced by cultural traditions (Pieniak et al., 2009), previous eating experiences (Papies, 2013), media/advertisements (Whalen et al., 2018), background nutrition knowledge (Miller and Cassady, 2015), choice context (Downs et al., 2015), product category (Plasek et al., 2020), packaging (Reutner et al., 2015), and health halo effects of labels such as "organic" (Perkovic and Orquin, 2018; Schuldt and Schwarz, 2010).

These factors contribute to the diverse and multidimensional knowledge representations that decision-makers draw upon when making food-related judgments and choices. Indeed, specific knowledge representations that are retrieved from memory (Scheibehenne et al., 2007) or explicitly provided to the decision-maker (Schulte-Mecklenbeck et al., 2013) can be used to make choices between food items using simple heuristics. Whereas knowledge representations may explain why people think some foods are healthier than others, they may be biased, causing systematic and predictable errors in healthiness perception. This could explain why people's judgments of healthiness deviate from an estimate of healthiness based on nutrient and energy values of the food (Orquin, 2014).

Researchers studying food judgment and choice typically rely on knowledge representations that are restricted to a predefined and limited set of factors and attributes (Steptoe et al., 1995). This also means that current approaches are not well suited for making generalizable predictions about judgments in the presence of interventions, such as different food labeling strategies (Kanter et al., 2018). How can we identify and quantify knowledge representations that underpin people's judgment of food healthiness? We propose a novel approach to overcome these challenges, which relies on recent advances in computational linguistics. Unlike previous approaches, in which food representations were either manually specified by the researchers or based on self-reports, we establish food representations using natural language data. More specifically, we use word distribution statistics in large text corpora to uncover quantitative representations for words and phrases that describe food items. The use of this type of data means that uncovered representations reflect information conveyed in language, which individuals may use to form beliefs, and may even guide everyday health judgment. We find support for this prediction by studying how knowledge representations retrieved from natural language can account for judgments of food healthiness across six experiments. Our further analysis reveals that our models perform well because they capture associations related to naturalness or rawness of foods.

The knowledge representations used in our analysis are high-dimensional vectors for words (also known as word embeddings) (Landauer and Dumais, 1997; Lenci, 2018; Mikolov et al., 2013). A useful property of word vectors is that the proximities between vectors measure the associations between individual words. These associations have been shown to correlate with human semantic, factual, probability, and social judgments (Bhatia, 2017; Caliskan et al., 2017; Pereira et al., 2016). Recently, researchers have shown that these word vectors can be used to quantify people's knowledge about various natural entities by using these as inputs into regressions that predict more complex (potentially non-associative) judgments in other domains (Bhatia, 2019; Richie et al., 2019; Zou and Bhatia, 2021).

Our approach is as follows: First, we obtain high-dimensional vector representations for food items from popular word embedding models trained on large-scale textual datasets. We hypothesize that these word vectors may serve as a good approximation of knowledge representations that underpin judgments of food healthiness. To test this proposition, with some training data involving people's ratings of diverse food items, we learn a mapping from our high-dimensional vector space to the (one-dimensional) scale that measures perceptions of healthiness (i.e. people's judgments). We then apply this mapping to food items outside of the training data to predict people's judgments for these "out-of-sample" food items. Note that such a mapping identifies regions of the vector space implicitly associated with food healthiness, and thus can be used to understand the conceptual and associative underpinnings of health judgment. We can also build this kind of mapping separately for different groups of people, to predict judgments of both lay and expert judges, as well as differences in judgments between individuals exposed to different front-of-pack labeling strategies. Across six studies, we demonstrate the generalizability, accuracy, and power of our approach.

## 3.3 Studies 1A, 1B, and 1C

Our primary objective was to establish the feasibility of our computational approach in predicting people's judgments of food healthiness. Therefore, we elicited judgments of healthiness for a wide range of food items (presented as food names) from the general population (Study 1A) and from a sample of registered dietitians (Study 1B). In Study 1C, we tested the performance of our models on healthiness judgments of foods' names and images.

### 3.3.1 Methods

#### Participants

Our approach does not rely on standard null hypothesis testing but rather on maximizing out-of-sample predictions. Using previous work for guidance (Bhatia, 2019), we chose to obtain judgments for a diverse set of 172 foods and aimed to recruit at least 100 participants (with each participant judging each of the 172 food items). The only exception was in Study 1B where we prioritized how many responses we could obtain from nutritional experts in a three-month window. Note that in all studies, the primary unit of analysis was the average healthiness rating, across all participants, for a given food item.

In all studies, only participants over 18 years of age were eligible to take part. Our only exclusion criterion was based on the correlation between each person's food ratings and the grand mean of aggregate ratings for those foods within the sample. Prior to data analysis (in all studies reported in this paper), we removed participants with a correlation lower than 0.4 with the grand mean of all ratings in a given study (based on the inspection of data from Study 1A). Although this exclusion criterion aimed to remove participants with very noisy ratings that would generate outlier responses, an analysis of the full sample shows that none of our results are affected by this exclusion criteria (see Section 3.6.1 of the Supplementary Materials).

For Study 1A, 149 participants were recruited from Prolific Academic in return for a fixed payment of £1.30. Using the aforementioned criterion, data from 15 respondents were removed leaving 134 participants as our final sample (aged 18-74 years, $M_{age}$ = 29.57 years, SD = 8.86, 43% females, and 84% had no dietary restrictions). For Study 1B, we contacted registered dietitians after a formal introduction by email with a request to take part in our study and forward the invitation to their colleagues. We also advertised the study on personal social media accounts. As an incentive, participants were able to request a report of the main findings. Nineteen registered dietitians took part in the study (none excluded, aged 23-56, $M_{age}$ = 35.84 years, SD = 10.36, 89% females and 68% had no dietary restrictions). One hundred participants recruited on Prolific Academic took part in Study 1C in return for a fixed payment of £1.90. We excluded one participant based on the same criteria as in Study 1A. This left 99 participants in our final sample (aged 18-69 years, $M_{age}$ = 27.25 years, SD = 10.20, 44% females, and 82% had no dietary restrictions). A detailed breakdown of participants' characteristics for this and other studies reported here is provided in the Supplementary Materials (Section 3.6.2). This research was approved by the University of Warwick's Biomedical and Scientific Research Ethics Sub-Committee (approval # REGO-2018-2268).

### Design and Procedure

In all studies, participants were asked to simply judge the healthiness of 172 foods on a scale ranging from -100 (extremely unhealthy) to +100 (extremely healthy). In Study 1A and 1B, each food was described using its name only. In Study 1C, a generic image of the food item was presented directly below the food name. Responses were made using a slider, with its starting position always at zero by default (neither healthy nor unhealthy). This scale was chosen because it is fine-grained (200 intervals) and balanced (symmetric around 0), offering nearly continuous data for predictive modeling (Bhatia, 2019). Participants had the option of selecting "Don't know" if they were unfamiliar with a food item, with those ratings removed from the analysis. The order of the items was randomized 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. After rating all foods, participants were asked about their age, gender, and dietary restrictions (with the options of "Pescetarian (no meat, but eat fish and/or shellfish)", "Vegetarian", "Vegan", "Other (please specify if you wish)" and "None of the above"). Our nutritional experts in Study 1B were also asked two additional demographic questions at the end of the survey (namely, "No. of Years as a Registered Dietitian" and "Area of Specialism").

### Materials

We obtained a list of foods from the USDA Food Composition Database, the most recent official publication of nutrient information pertaining to over 3102 unique food items (U.S. Department of Agriculture, 2019). Only foods present in the vocabulary of the pre-trained word2vec model were considered, leaving 571 food items (see the Computational Approach section for detail). Two hundred food items, across all food categories (e.g. vegetables, meats, dishes), were manually chosen by co-author WZ to ensure diversity in the stimuli set. Next, co-authors NG and LW removed uncommon and ambiguous food items such as "squash" (because of its additional meaning related to sports), resulting in the final list of 172 food items (see the OSF repository associated with this project for the full list: https://osf.io/jys6u/). Note that the same list of 172 food items was used in all studies reported in the main text of the manuscript.

In Study 1C, 69 of the food images were directly sourced from an image database for experimental research (Blechert et al., 2019), with the remaining 103 images sourced online and standardized to match (white background, 600 x 450 dimensions, and jpeg format).

**Computational approach**

We used three statistical models to predict subjective food healthiness judgments. Our analysis relied on participants' judgments at the aggregate level. We evaluated the accuracy of each of our three statistical models in predicting subjective food healthiness judgments using leave-one-out cross-validation, which means that we trained our models on all but one aggregate judgment ("training data") and used the trained model to predict the rating of the left-out food item ("test data"). We repeated this procedure 172 times so that each food item was in the test data once. Cross-validation ensures that our modeling avoided overfitting and that performance of each model was evaluated based on model generalizability.

In the first model, the Nutrient Model, we used nutrient content information to predict healthiness judgments. This model was an ordinary-least-squared regression with main effects for food calorie content, amounts of nutrients (fat, saturates, sugar, salt, and protein) per 100g, and the relative coding scheme based on the UK traffic light labeling for fat, saturates, sugar and salt (green, amber and red). Under the traffic light labeling system, green signifies a healthier food choice to consumers implying "go ahead"; amber indicates the item contains moderate amounts of the negative nutrient(s); and red signals caution for overconsumption (Trudel et al., 2015). The model was fit on the training data, and the best fitting parameters of the model were applied to the nutrient information of the (out-of-sample) food, in order to predict participant ratings. The nutrients and calorie information included in the Nutrient Model reflects the current European Union's regulations concerning mandatory information for food package labeling (Article 30, Regulation No. 1169/2011 European Commission, 2011). In the Supplementary Materials (Section 3.6.3), we summarize tests of the robustness of our results using three extended versions of the Nutrient Model. First, we expanded the Nutrient Model to incorporate the potential role of 23 nutrients (e.g., fiber, calcium, and Vitamin C). Second, we also tested a version of the model that used nutrient amounts per portion size, defined as the amount per 100 calories. We also combined these two extensions into our final, third model.

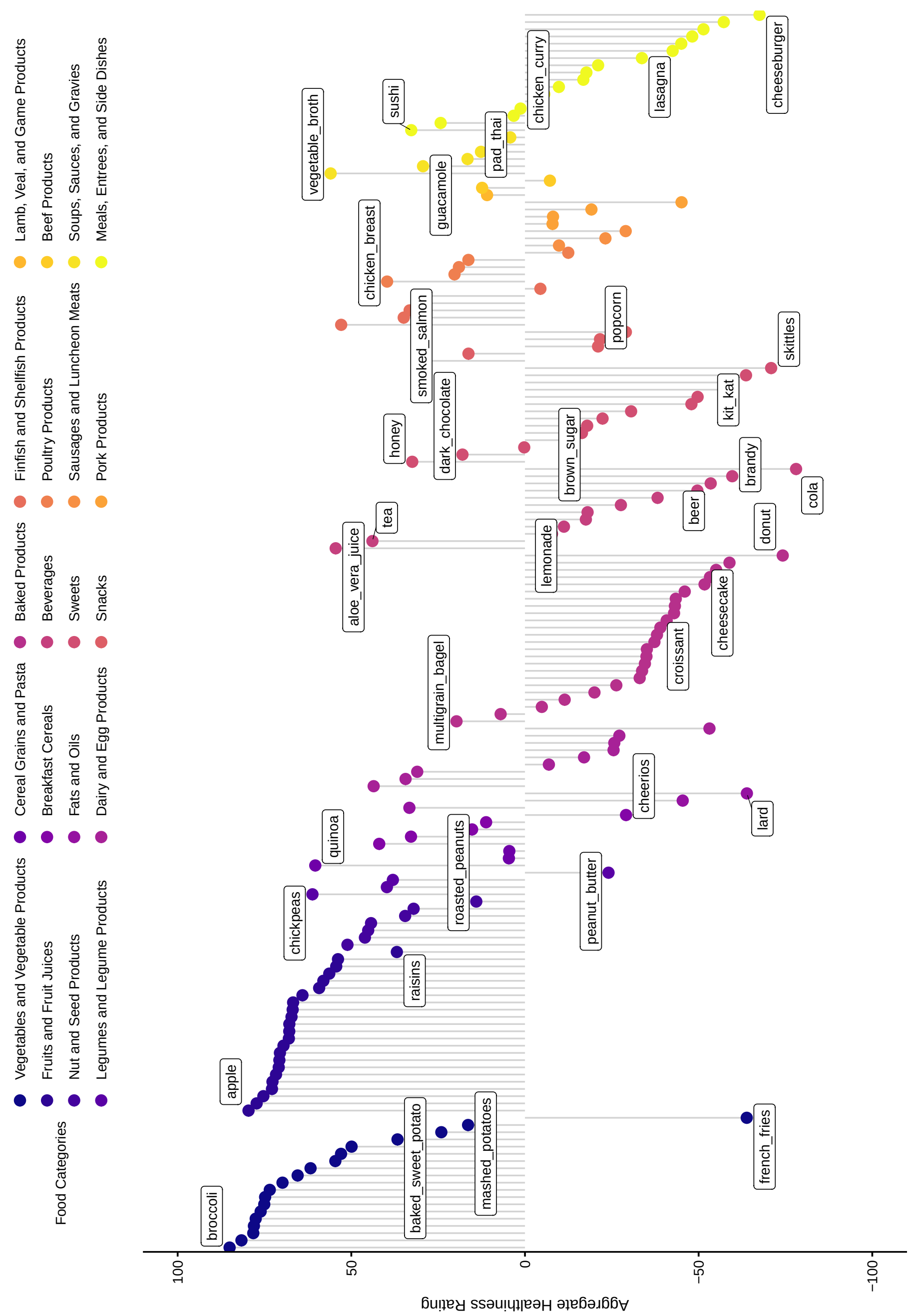
In our Vector Representation Model, we used vector representations from the word2vec model (Mikolov et al., 2013). This model was pre-trained on a dataset of Google News articles, and has 300-dimensional vector representations for three million common words and phrases in the English language (see Mikolov et al., 2013 for details). In designing our studies, we only considered foods whose name features in the pre-trained word2vec model. We also analyzed the predictions of other established pre-trained word vector models, which we discuss in the Supplementary Materials (Section 3.6.4). In our main analysis, we used normalized word vectors, in which the magnitude of the vectors was scaled to be equal to 1. We regressed participants' healthiness ratings on these vectors, which allowed the model to learn a linear mapping from the semantic vector space to health judgments. This learned mapping was then applied to the vectors of other (out-of-sample) foods, in order to predict participant ratings of those foods, and measure the models' predictive accuracy. Because of the high number of predictor variables in this model (300), we applied a regularized 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 previous and similar work, ridge regression was the best-fitting regression technique for mapping pre-trained 300-dimensional vector representations to judgments and was consequently chosen for our analysis (Bhatia, 2019; Richie et al., 2019). We also tested other regression techniques including lasso, support vector, and k-nearest neighbors regression and found ridge regression was indeed the best-fitting regression. We discuss this robustness test in the Supplementary Materials (Section 3.6.5). Finally, our third Combined Model concatenated the 11-dimensional Nutrient Model with the 300-dimensional Vector Representation Model. Using ridge regression, we explore the extent that both models can collectively explain people's subjective food healthiness judgments.

### 3.3.2 Results

We began by examining the distribution of aggregate healthiness ratings in Figure 3.1. Here we observed that healthiness judgments varied greatly amongst 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 77 and 82 for tomatoes, cucumber, apple, carrots, and broccoli, respectively. The five foods that received the unhealthiest ratings, ranging between -65 and -50, were cola, donut, skittles, cheeseburger, and Kit Kat.

**Figure 3.1**

Distribution of aggregated food healthiness ratings from Study 1A

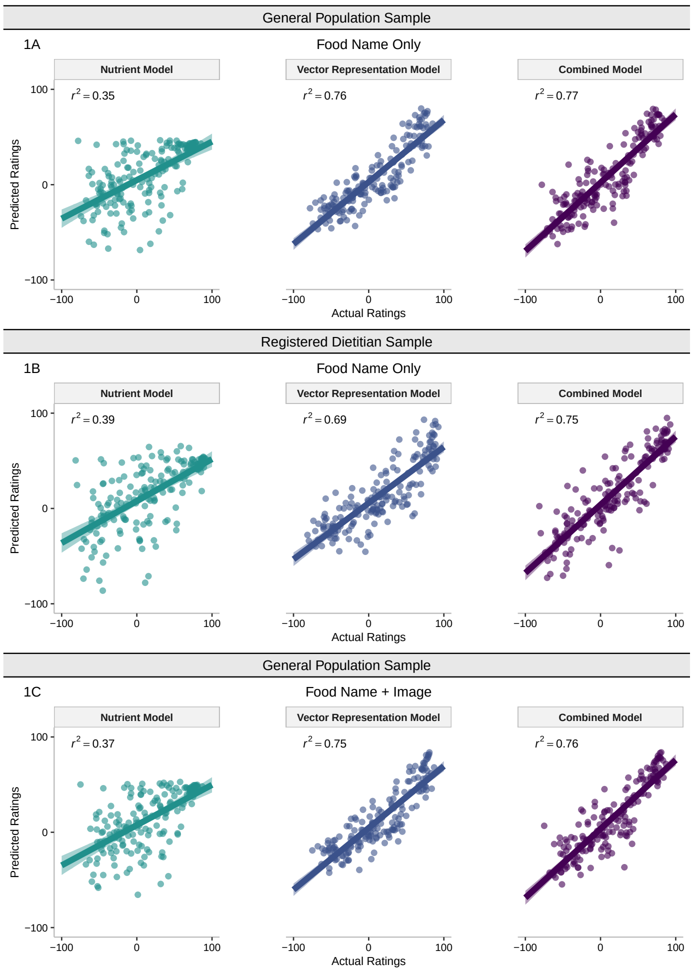



*Note.* For clarity, foods are separated by food category. Text labels indicate exemplar food items in each of the categories.

Figure 3.2 summarizes the accuracy rates of our three models in Studies 1A, 1B, and 1C. The dots within each scatterplot represent the out-of-sample predicted vs actual (aggregated) healthiness ratings for the individual foods. As we are using predictive modeling, the coefficient of determination ($r^2$) reflects the performance of the model when making out-of-sample predictions.

**Figure 3.2**

A Comparison of Predictive Accuracy between Models in Studies 1A, 1B, and 1C




*Note.* A comparison of predictive accuracy between models that used only nutrient content, only word vector representations, or a combination of nutrient content and word vector representations in a general population sample (1A), expert sample (1B) and with food images included as stimuli (1C).

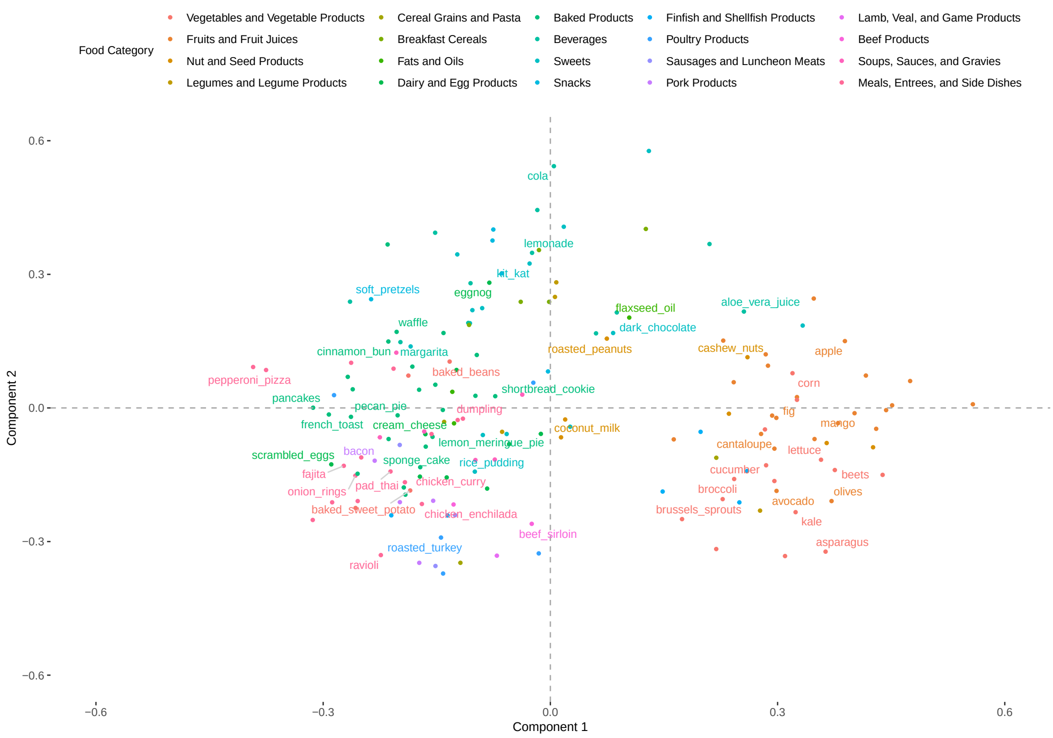
As shown in Figure 3.2, the out-of-sample predictive accuracy of the Vector Representation Model was very high across all studies, with an $r^2$ ranging from 0.69 (95% CI [0.63, 0.75] to 0.76 (95% CI [0.71, 0.81]). By comparison, the predictive accuracy of the model based on the foods' nutritional information was always lower ($r^2$ ranging from 0.35, 95% CI [0.24, 0.46] to 0.39, 95% CI [0.28, 0.50]). The Combined Model performs best however, achieving marginally higher predictive accuracy than the Vector Representation Model in every study ($r^2$ ranging from 0.75 (95% CI [0.70, 0.80] to 0.77, 95% CI [0.72, 0.82]). Overall, these findings highlight that the performance of the Vector Representation Model is stable, even when using ratings from participants with high nutritional expertise and with food images as stimuli. We performed several robustness checks to assure the reliability of our findings. First, we ran separate paired sample t-tests to compare the squared errors from different models for each study (see Section 3.6.6 in the Supplementary Materials). Across all studies, the mean squared errors from the Vector Representation Model and the Combined Model were significantly lower than those from the Nutrient Model (all $p < 0.01$). We also repeated our analysis at the individual level, without aggregating healthiness ratings for each food. Results are presented in our Supplementary Materials (Section 3.6.7) and show that our findings remain largely unchanged. Section 3.6.4 of the Supplement summarizes $r^2$ for the Vector Representation Model based on alternative word vectors obtained from fastText (Mikolov et al., 2017) and GloVe (Pennington et al., 2014). Finally, in Section 3.6.5, we show the results of different regression techniques, including lasso, support vector, and k-nearest neighbors. Once again, using alternative word vectors or regression techniques did not alter our results.

Returning to the results from the Vector Representation Model based on the ridge regression and word2vec vectors, our approach was also able to capture qualitative trends in our data by correctly predicting the categories of foods judged as being high or low in healthiness. For example, both observed and predicted ratings were highest for categories such as Fruits and Fruit Juices, Vegetables and Vegetable Products, and Nut and Seed Products. Likewise, both observed and predicted ratings were lowest for categories such as Baked Products, Sweets, and Fats and Oils. In fact, when pooling the data by food category, we found the Vector Representation Model predicted average healthiness ratings for categories of foods with an out-of-sample $r^2$ of 0.83 (95% CI [0.79, 0.86]). The Nutrient Model, in contrast, achieved an $r^2$ of only 0.31 (95% CI [0.20, 0.41]). It seems healthiness judgments are sensitive to the category of the food item, a property easily captured by the Vector Representation Model, but less so for the Nutrient Model (Orquin, 2014). Further details of this analysis are provided in Section 3.6.8 of the Supplementary Materials.

The reason why the Vector Representation Model performs well is that it may capture the latent associations underpinning judgments of healthiness. To explore these associations, we applied a Principal Component Analysis to the vector representations of the 172 food items. Projections for the first two components are shown in Figure 3.3. By inspecting Component 1, it is clear that negative values correspond to mostly heavily processed and junk foods (e.g., pepperoni pizza, bacon, onion rings), whereas positive values correspond mainly to organic and unprocessed vegetables and fruits (e.g. apple, mango, lettuce, beets). Component 2 on the other hand, appears to reflect the sweetness/sugariness of the food. The most positive scoring foods on this component are sugary drinks (e.g. cola, lemonade) and sugary snacks (e.g. kit kat, dark chocolate). Among the negative scores for Component 2, we can see meats (e.g., roasted turkey) but also less sugary vegetables (e.g., brussels sprouts, asparagus).

**Figure 3.3**

A 2D projection (based on the Principal Component Analysis) of vector representations for the 172 food names

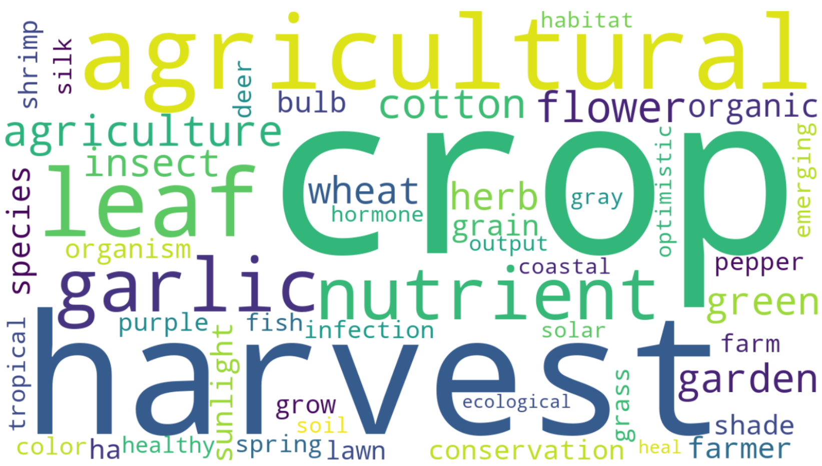



*Note.* For clarity, only a random subset of 50 names are labeled on the plot.

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 3.4 shows a word cloud of the fifty English language words with the highest healthiness predictions, derived with this approach (from a model trained on Study 1A ratings). See Appendix A for words clouds trained on ratings from all other studies. 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/degree of processing, organic, appearance) identified by previous researchers as being psychological cues for food healthiness (Siipi, 2013). Our novel computational approach provides quantitative methods for uncovering these associations.

**Figure 3.4**

Healthiest "Other" Words based on the Vector Representation Model




*Note.* Words with the highest predicted healthiness ratings based on the Vector Representation Model estimated for the lay people in Study 1A.

## 3.4 Studies 2A, 2B and 2C

Studies 1A-1C are based on judgments of food names and images. Yet, when people evaluate foods, they are typically provided with additional nutritional information. Moreover, many food labeling policies rest on the assumption that consumers make more informed decisions if nutritional information is prominently displayed on food packaging. The relatively poor performance of the Nutrient Model documented in Studies 1A-1C may reflect people's lack of awareness (or memory) of the nutritional values of the individual food items. Accordingly, if the nutritional information was made more salient with the use of real food labeling strategies, we might expect that the contribution of nutrients in the Nutrient Model would increase relative to the Vector Representation Model. Finally, the results of Studies 1A-1C uncovered underlying associations between foods and naturalness, or rawness, which underpin food healthiness judgments. However, it is important to determine whether these associations continue to play a role even if foods' nutritional values are made more salient.

We addressed these issues in Studies 2A-2C by eliciting food healthiness judgments from participants who saw either food names alone (as in Study 1A) or food names along with the label highlighting various aspects of its nutrition. Again, we recruited adult participants for this series of studies. In Study 2A, we provided our treatment group with information about the calories per 100g. The provision of calorie information to aid healthy eating is supported by qualitative research showing that consumers use energy content information (calories) as a proxy for the overall nutritional value of a product (Van Kleef et al., 2008). In Studies 2B and 2C, we examined the effects of information about key nutrients (fat, saturates, sugars, and salt). Under EU rules, front-of-pack labeling of this kind is acceptable with either no color-coding or traffic-light colored cues (i.e., red highlights high, orange medium, and green low amounts of fat, saturates, sugar, and salt) (European Commission, 2011). While both strategies highlight individual nutrients, it is the color-coded format that also aids consumers to judge whether a particular amount is high, medium, or low. In Study 2B we gave the treatment group nutrient labels without color-coding and in Study 2C we gave this group with nutrient labels with color-coding.

### 3.4.1 Methods

#### Participants

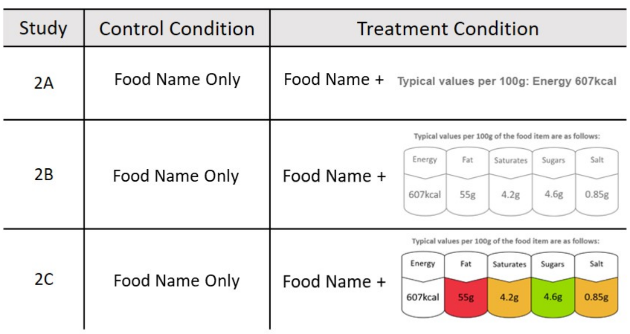
There were 202 participants in Study 2A, and after the removal of five using our exclusion criteria, 197 participants were included in the final analysis (aged 18-71 years, $M_{age}$ = 30.30 years, SD = 10.74, 52% female, and 80% had no dietary restrictions). From the initial 199 participants who took part in Study 2B, four were excluded leaving 195 participants (aged 18-65 years, $M_{age}$ = 29.16 years, SD = 10.28, 48% female, and 82% had no dietary restrictions). Finally, 202 participants took part in Study 2C (aged 18-78 years, $M_{age}$ = 34.69 years, SD = 11.51, 70% female, and 81% had no dietary restrictions). No participants were excluded from this study as all participant responses fell above the threshold for removal. Only residents of the UK were allowed to participate in Study 2C to assure knowledge and familiarity with the traffic light food labeling system.

### Design and Procedure

We tested the role of food labeling on judgment, in which we gradually (across studies) introduced more informative (and realistic) formats of food labeling. All three studies used a between-subjects design. In half of the sample (control group) participants rated food healthiness of 172 foods in the same manner as in Study 1A and 1B. In the treatment groups, participants rated each of the food names presented alongside a nutrition label. In Study 2A, this was the energy (kcal) amounts per 100 grams of the food. In Study 2B, we additionally included the absolute amount of fats, saturates, sugars and salts. Finally, in Study 2C, we used the same objective information as above, but also added the "traffic light" system used in the UK, which indicates the relative amount of different nutrients, categorizing them into green, amber, and red groups. The examples of the labeling used in each study are presented in Figure 3.5.

**Figure 3.5**

Food stimuli presented to participants in Studies 2A, 2B and 2C




*Note.* All participants were from a general population sample.

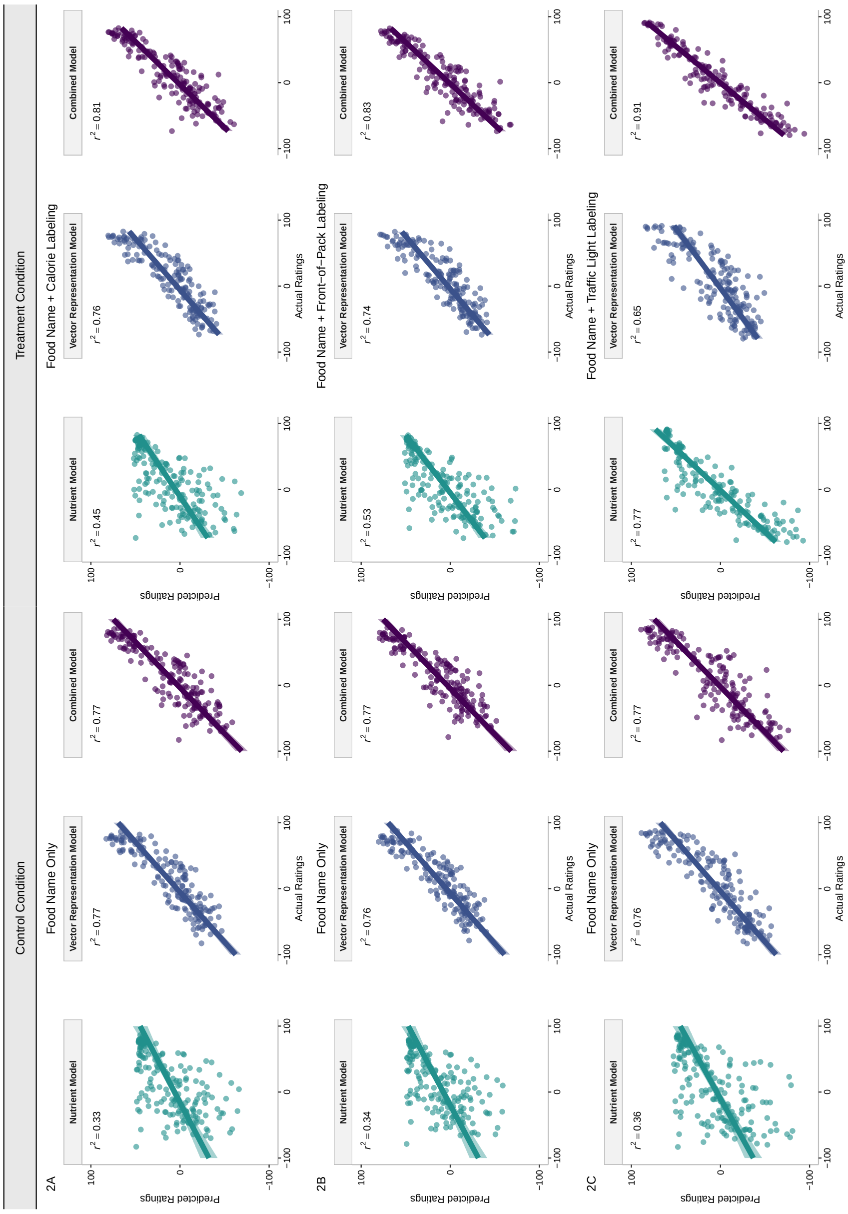
### 3.4.2 Results

As shown in Figure 3.6, the Vector Representation Model performed very well across all studies and conditions. In fact, the out-of-sample predictive accuracy of the Vector Representation Model was very high, with $r^2$ ranging from 0.65 (95% CI [0.59, 0.72]) to 0.77 (95% CI [0.72, 0.81]), in each study and condition. By comparison, the predictive accuracy of the models based on the foods' nutritional information was lower but also much more variable ($r^2$ ranging from 0.33, 95% CI [0.22, 0.44] to 0.77, 95% CI [0.72, 0.83]). Figure 3.6 reveals a systematic pattern – the predictive accuracy of the Nutrient Model increased with the amount of the nutritional information presented alongside foods' names. This is unsurprising as it shows that people integrated label information into their judgments (González-Vallejo et al., 2016; Scarborough et al., 2007a). Despite this, the Vector Representation Model still performed better than the Nutrient Model when participants saw only calorie information (Study 2A) and calorie information with front-of-pack nutrient labeling (Study 2B). Only in the most informative condition, traffic light labeling (Study 2C), did the Nutrient Model outperform the Vector Representation Model. Figure 3.6 also shows that the accuracy of the vector representation model is identical across the two conditions in Studies 2A and 2B, although it does drop slightly in Study 2C. This is not significant, as can be seen from the slight overlap in 95% CIs of the control ($r^2$ = 0.76, 95% CI = [0.71, 0.81]) and traffic light labeling conditions ($r^2$ = 0.65, 95% CI = [0.59, 0.72]) in Figure 3.10 of the Supplementary Materials. In any case, these results show that associations with food names play an important role in people's judgments of healthiness, often more than its nutritional composition.

Figure 3.6 also summarizes the predictive accuracy of the Combined Model – which uses both the word vectors and nutritional information to predict people's judgments. In 5 out of 6 cases, the Combined Model performed better than the individual models. The highest accuracy was achieved in the traffic light labeling condition, with $r^2$ of 0.91 (95% CI [0.89, 0.93]), which was markedly higher than 0.77 (95% CI [0.72, 0.83]) of the Nutrient Model and 0.65 (95% CI [0.59, 0.72]) 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.

**Figure 3.6**

A Comparison of Predictive Accuracy between Models in Studies 2A, 2B and 2C

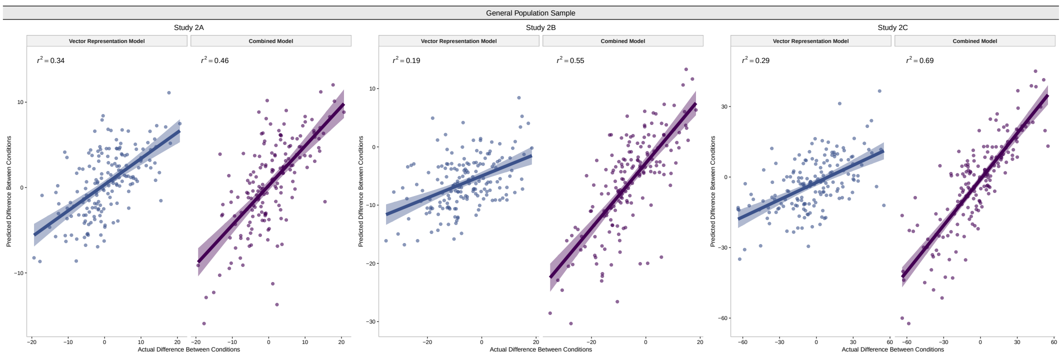



*Note.* A comparison of predictive accuracy between models that used only nutrient content, only word vector representations, or a combination of nutrient content and word vector representations. Actual ratings were all from a general population sample who were randomly assigned to either the control or the treatment condition in each study.

Do representations of foods change when nutrient information of foods is highlighted? In other words, do we observe a systematic shift in knowledge representations due to the various types of food labeling? To answer this question, we computed differences between aggregate ratings for each food made in the condition with and without a food label. We then refitted our Vector Representation and Combined Model with these difference scores as a dependent variable. Figure 3.7 shows that the Vector Representation Model explains a non-trivial amount of variance in the difference between the conditions in all three studies. At the same time, the predictive accuracy of the Combined Model increases markedly from Study 2A through to 2C, which confirms that people who saw additional nutritional information did in fact rely on it when making their judgments. This supports the interpretation that even if food labeling changes how people make judgments, the reliance on knowledge representations captured by the Vector Representation Model remains stable and influential. It also shows that word vector representations can predict the idiosyncratic effects of nutrient labels on health judgments for different food items.

**Figure 3.7**

Leave-one-out cross-validation results for the ability of vector representations to predict condition differences




*Note.* Leave-one-out cross-validation results for the ability of vector representations (Vector Representation Model) and the Combined Model to predict the difference between conditions in Study 2A (calorie information – control), Study 2B (calorie information with front-of-pack nutrient labeling – control) and, Study 2C (traffic light labeling – control).

## 3.5 General Discussion

Everyday dietary decisions are influenced by people's subjective perceptions of food healthiness. Psychological explanations of this process are incomplete without an accurate model of the rich knowledge and diverse associations underpinning people's judgments of what food is healthy and what is not. In this paper, we offer a novel method for uncovering these knowledge representations by combining insights from machine learning and computational linguistics. Using vector representations of food items derived from natural language, we show that it is possible to predict healthiness judgments highly accurately. We show that people's judgments can be partly explained by the strength of association between individual food items and concepts pertaining to naturalness (e.g., harvest, leaf) and rawness (e.g., crop, organic). These associations play a role even when judgments are made in the presence of food images or are made by trained dietitians. In addition, high accuracy rates obtained by our Combined Model indicate that such knowledge representations in language do not merely reflect beliefs about nutritional composition; rather they capture something unique about people's associations with different foods. Thus, our models can help evaluate how different front-of-pack labeling strategies influence food healthiness judgments.

Unlike previous approaches, our method does not require us to identify specific factors or attributes that we, as researchers, believe to be related to healthiness judgments. Instead, by using our best-fit model to predict the "healthiness" of common words in the English language, we show that nature-related words such as "crop", "harvest", and "agricultural" are implicit in people's judgments. This is consistent with other findings that perceived naturalness and healthiness are often intertwined (Sanchez-Siles et al., 2019; Siipi, 2012 but see Fernbach et al., 2019). These results also align with the finding that rawness or the degree to which a food has been processed is a strong cue of healthiness in food choice (Scheibehenne et al., 2007; Schulte-Mecklenbeck et al., 2013). Notably, our results indicate that models based on these associations are accurate even if participants are explicitly told about the nutritional composition of foods.

Our approach offers a unique insight into the psychological basis of subjective food healthiness judgments by exploring foods in their most abstract forms (name or image). That said, a model trained on written text is unlikely to accurately capture sensorimotor information about foods (e.g. smell, texture), which would also be relevant in real-world situations (De Deyne et al., 2016; Lynn et al., 2020; Papies et al., 2020). Hence, while our results are promising, they are only a first step in providing a rich set of attributes and associations that people use in judging foods' healthiness.

Neither explicit food labeling nor expert judgments reduced the contributions of the knowledge associations established by the Vector Representation model. With respect to expert judgments, these findings are in line with research showing that nutritional expertise does not always translate into a higher reliance on nutritional information when making healthiness judgments (Orquin, 2014). Our results also speak to the value of nutritional labeling more generally. Given that associations played a role in all studies, existing front-of-pack labeling can neither substitute nor correct for the associations that people rely on when judging foods' healthiness.

There are many potentially useful applications of our computational approach. Future studies could test the predictive ability of this Vector Representation Model with and against other formats of nutrient labeling such as France's Nutri-score label (color-coded without numerical information). Thus, the use of this approach could be vital in determining a single internationally agreed nutrient labeling system (Goiana-da Silva et al., 2019), especially since it provides directly comparable results between labeling formats. However, further work is necessary to establish whether the accuracy of our models changes when participants are presented with other information present on pre-packaged foods, such as branding, health claims, and back-of-pack nutrition labeling.

An important outstanding question is whether our Vector Representation Model is generalizable to judgments of other foods than the 172 items tested in all six studies. In Section 3.6.10 of our Supplementary Material, we report the results of a new study in which we elicited judgments of 60 new foods from a sample of 97 participants. Instead of training a new model, we used the Vector Representation Model from Study 1A to derive predictions for our new foods. Our models performed very well –with our approach we can predict healthiness judgments of new foods from a new group of participants highly accurately. To assist future research, we have obtained predictions of our models for hundreds of novel food items and made these available via OSF (https://osf.io/jys6u/). These can be used to evaluate future interventions and to test alternative psychological mechanisms that underpin human judgments and choices of foods. Overall, our studies provide new insights into people's food healthiness judgments, while our methods offer an exciting new avenue to researchers and practitioners interested in designing interventions for healthy eating.

## 3.6 Supplementary Materials

In this section, we perform a series of robustness checks to provide a wider context for the effectiveness of using a vector space approach to uncover insights about food healthiness perceptions. We also provide further details about our study to increase reproducibility.

**Section 3.6.1 - Full Sample Results:** This demonstrates that the exclusion of participants with seemingly irrational responses does not affect the performance of our models.

**Section 3.6.2 - Demographic Characteristics:** A breakdown of the participant demographic information, including the details of the experience and specialty of the registered dietitians in Study 1B to support future research in this area.

**Section 3.6.3 - Extended Nutrient and Combined Models:** Incorporates additional nutrient data and amounts per portion size, to assess whether considering these factors improves the performance of these comparative models.

**Section 3.6.4 - Alternative Word Embeddings:** Compares our chosen pre-trained word embedding model (word2vec) with two other popular pre-trained word embedding models that have slightly different training algorithms. To clarify, the purpose of this is not to compare word embedding approaches or advocate for a single model, but rather to assess the robustness and consistency of the broader vector-space-based approach.

**Section 3.6.5 - Secondary Vector Representation Models:** Details the different regression techniques used to determine the accuracy of our Vector Representation Model and Combined Model. The reason we replicated our tests using multiple machine learning regularization techniques was because we faced a high-dimensional regression problem (where we had over 300 predictor variables). As a result, we could not use standard linear regressions so chose to test techniques that have previously been found to address this problem, akin to other work using a similar computational approach (Bhatia, 2019; Bhatia et al., 2021a).

**Section 3.6.6 - Test for Model Comparison:** Explicit statistical comparisons of our computational models using separate paired sample t-tests.

**Section 3.6.7 - Individual-level Modeling:** Model performances in each condition based on the individual's set of judgments, to ensure our findings still hold at the individual level.

**Section 3.6.8 - Food Category Modeling:** Model performances in each condition when foods are aggregated to food category level.

**Section 3.6.9 - Word Cloud for Unhealthy Associations:** The 50 non-food item words in American English that are given the lowest healthiness ratings by our Vector Representation Model, trained on participant ratings from Study 1A.

**Section 3.6.10 - Test of Model Generalizability:** A study to determine the accuracy of the Vector Representation Model healthiness rating predictions for food names not used to train the model.

### 3.6.1 Full Sample Results

In our main manuscript, we excluded participants whose ratings correlated poorly ($r < 0.4$) with average participant ratings. To make sure that the exclusion criterion did not significantly alter the predictive accuracy of our models, we also fit the three main models to the original data without the exclusion. Table 3.1 reports the out-of-sample $r^2$ with either 95% CI for the Nutrient Model, Vector Representation Model, and Combined Model trained on average healthiness ratings. Compared to the results reported in Table 3.6 (see Test for Model Comparison Section below), model performance stayed the same when including all participants.

**Table 3.1** Aggregate level out-of-sample $r^2$ for the Vector Representation Model, Nutrient Model and Combined Model fit to the original data including all participants, numbers in the brackets represent 95% CIs

| Study | Nutrient Model | Vector Representation Model | Combined Model |
|---|---|---|---|
| 1A | 0.35 [0.24, 0.46] | 0.76 [0.71, 0.80] | 0.76 [0.71, 0.82] |
| 1B | 0.39 [0.28, 0.50] | 0.69 [0.63, 0.75] | 0.75 [0.70, 0.80] |
| 1C | 0.37 [0.26, 0.48] | 0.76 [0.71, 0.80] | 0.77 [0.72, 0.82] |
| $2A_{\_control}$ | 0.33 [0.22, 0.44] | 0.77 [0.72, 0.81] | 0.77 [0.72, 0.82] |
| $2A_{\_treatment}$ | 0.45 [0.35, 0.56] | 0.76 [0.71, 0.81] | 0.81 [0.77, 0.85] |
| $2B_{\_control}$ | 0.34 [0.23, 0.45] | 0.76 [0.71, 0.81] | 0.77 [0.72, 0.82] |
| $2B_{\_treatment}$ | 0.53 [0.44, 0.63] | 0.74 [0.69, 0.79] | 0.82 [0.79, 0.86] |
| $2C_{\_control}$ | 0.36 [0.25, 0.47] | 0.76 [0.71, 0.81] | 0.77 [0.72, 0.82] |
| $2C_{\_treatment}$ | 0.77 [0.72, 0.83] | 0.65 [0.59, 0.72] | 0.91 [0.89, 0.93] |

### 3.6.2 Demographic Characteristics

**Table 3.2** Demographic Characteristics of all Participants

| Characteristic | Study 1A | | Study 1B | | Study 1C | | Study 2A | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Control (Lay) | | Control (Expert) | | Food Images (Lay) | | Control (Lay) | | Calorie Labeling (Lay) | |
| Age | | | | | | | | | | |
| *N* | 134 | | 19 | | 99 | | 96 | | 101 | |
| Mean | 29.57 | | 35.84 | | 27.25 | | 31.64 | | 28.96 | |
| SD | 8.86 | | 10.36 | | 10.20 | | 11.66 | | 9.81 | |
| Min - Max | 18 - 74 | | 23 - 56 | | 18 - 69 | | 18 - 71 | | 18 - 60 | |
| | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % |
| Gender | | | | | | | | | | |
| Female | 58 | 43 | 17 | 89 | 44 | 44 | 50 | 52 | 53 | 52 |
| Male | 73 | 54 | 2 | 11 | 54 | 55 | 45 | 47 | 47 | 47 |
| Other | 2 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| Prefer not to say | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Diet Restrictions | | | | | | | | | | |
| None | 113 | 84 | 13 | 68 | 81 | 82 | 71 | 74 | 86 | 85 |
| Pescetarian | 1 | 1 | 2 | 11 | 3 | 3 | 7 | 7 | 4 | 4 |
| Vegetarian | 9 | 7 | 1 | 5 | 6 | 6 | 3 | 3 | 3 | 3 |
| Vegan | 2 | 1 | 0 | 0 | 2 | 2 | 3 | 3 | 1 | 1 |
| Other | 9 | 7 | 3 | 16 | 7 | 7 | 12 | 13 | 7 | 7 |

**Table 3.2** Demographic Characteristics of all Participants (*Continued*)

| Characteristic | Study 2B | | | | Study 2C | | | | Supplement | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Control (Lay) | | Front of Pack Labeling (Lay) | | Control (Lay) | | Traffic Light Labeling (Lay) | | New Food Stimuli (Lay) | |
| Age | | | | | | | | | | |
| *N* | 104 | | 91 | | 102 | | 100 | | 97 | |
| Mean | 29.97 | | 28.35 | | 35.18 | | 34.19 | | 25.71 | |
| SD | 11.27 | | 9.28 | | 10.60 | | 12.41 | | 8.97 | |
| Min - Max | 18 - 65 | | 18 - 57 | | 18 - 64 | | 18 - 78 | | 18 - 60 | |
| | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % |
| Gender | | | | | | | | | | |
| Female | 42 | 40 | 51 | 56 | 77 | 75 | 64 | 64 | 38 | 39 |
| Male | 62 | 60 | 38 | 42 | 25 | 25 | 36 | 36 | 58 | 60 |
| Other | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| Prefer not to say | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Diet Restrictions | | | | | | | | | | |
| None | 85 | 82 | 75 | 82 | 82 | 80 | 82 | 82 | 85 | 88 |
| Pescetarian | 2 | 2 | 4 | 4 | 4 | 4 | 3 | 3 | 2 | 2 |
| Vegetarian | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 4 | 5 | 5 |
| Vegan | 4 | 4 | 2 | 2 | 4 | 4 | 5 | 5 | 2 | 2 |
| Other | 9 | 9 | 6 | 7 | 7 | 7 | 6 | 6 | 3 | 3 |

**Table 3.3** Demographic Characteristics specific to Nutritional Experts

| Characteristic | Study 1B | |
|---|---|---|
| | Control (Expert) | |
| | $n$ | % |
| No. of Years as a Registered Dietitians | | |
| Less than 1 year | 1 | 5 |
| 1-4 years | 8 | 42 |
| 5-9 years | 3 | 16 |
| 10-19 years | 3 | 16 |
| 20 years or more | 4 | 21 |
| Area of Specialism | | |
| Diabetes | 3 | 16 |
| Gastroenterology | 4 | 21 |
| Older People | 1 | 5 |
| Oncology | 1 | 5 |
| Pediatric | 2 | 11 |
| Parenteral and Enteral Nutrition | 1 | 5 |
| Renal Nutrition | 6 | 32 |
| Other | 1 | 5 |

### 3.6.3 Extended Nutrient and Combined Models

In testing the accuracy of the Nutrient Model our assumption was that participants' healthiness ratings may be best reflected in people's knowledge of the nutritional/energy composition of individual foods. In our main manuscript, our Nutrient Model included information about calories and nutrients (fat, saturates, sugar, salt, and protein) per 100g, as well as the traffic light color-coding. Whereas this approach reflects current EU regulation about food labeling, it may not truly reflect the nutritional information that people actually rely on when making their judgments. In the following section, we provide further robustness checks to the results we report in the main manuscript by using three different extended versions of the Nutrient Model. First, most of the existing regulation is focused on highlighting nutrients that are typically associated with weight gain and poorer health. We, therefore, extended our original Nutrient Model by adding (up to) 23 positive nutrient variables (including fiber, calcium, and vitamin C). Second, we consider the possibility that perceptions of healthiness may better align with nutritional information expressed in relation to the portion size amounts of each food. As there is no agreement of what portion size should be, we made a judgment call to use nutrient amounts per 100 calories (thus g/100kcal). This way, our Nutrient Model does not punish foods that would never be eaten in large volumes (e.g., chewing gum).

Since many of the foods do not have all the micronutrients, we used ridge regression instead of linear (ordinary least square) regression for the two alternative Nutrient Models and Combined Models with micronutrients added. Further analysis also confirmed that ridge regression is more appropriate for the extended Nutrient Model as it achieved higher predictive accuracy than linear regression.

As can be seen in Tables 3.4 and 3.5, the use of portion size amounts does not improve the accuracy of either the Nutrient Model or the Combined Model in any condition or study. The addition of positive nutrients does not improve the accuracy of the Combined Model but impairs the accuracy of the Nutrient Model. The extended Nutrient Model with the added micronutrients (either per 100g or per 100kcal) performed considerably worse than the original Nutrient Model. The predictive power of the extended Nutrient Model was highly variable ($r^2$ ranging between 0.07 (95% CI [0.01, 0.14]) to 0.72 (95% CI [0.66, 0.78]), increasing across the conditions where nutritional information was accessible to participants and performed best in the traffic light labeling condition of Study 2C. Similarly, the alternative Combined Models performed consistently with the original Combined Model. In all control conditions and Studies 1A and 1C, the $r^2$ of the extended Combined Model with micronutrients (either per 100g or per 100kcal) varied between 0.65 (95% CI [0.57, 0.73]) and 0.89 (95% CI [0.86, 0.92]). It performed slightly worse in the sample of registered dietitians and reached the highest out-of-sample accuracy in the treatment condition of Study 2C. Moreover, using separate paired-sample t-tests for each study, we found that the mean squared errors from the extended versions of the Nutrient Model were not significantly different from those of the original Nutrient Model. This result holds for the Combined Model, suggesting that adding more nuanced nutrient information did not improve the predictive accuracy of the Nutrient Model or the Combined Model. The detailed statistics are available upon request. Taken together, these robustness checks show that the relatively lower accuracy of the Nutrient Model was not due to the incorrect assumption about the role of micronutrients or portion size considerations in people's judgments.

**Table 3.4** Out-of-sample $r^2$ comparisons between the original and alternative Nutrient Models

| Study | Nutrient Model (per 100g) | Nutrient Model (per 100 kcal) | Nutrient Model with micronutrients (per 100g) | Nutrient Model with micronutrients (per 100 kcal) |
|---|---|---|---|---|
| 1A | 0.35 [0.24, 0.46] | 0.34 [0.23, 0.45] | 0.12 [0.05, 0.20] | 0.12 [0.05, 0.19] |
| 1B | 0.39 [0.28, 0.50] | 0.38 [0.27, 0.49] | 0.15 [0.07, 0.23] | 0.14 [0.06, 0.22] |
| 1C | 0.37 [0.26, 0.48] | 0.33 [0.23, 0.44] | 0.18 [0.09, 0.26] | 0.11 [0.04, 0.18] |
| 2A$_{_control}$ | 0.33 [0.22, 0.44] | 0.32 [0.21, 0.43] | 0.08 [0.02, 0.14] | 0.07 [0.01, 0.14] |
| 2A$_{_treatment}$ | 0.45 [0.35, 0.56] | 0.42 [0.31, 0.52] | 0.26 [0.17, 0.35] | 0.24 [0.15, 0.33] |
| 2B$_{_control}$ | 0.34 [0.23, 0.45] | 0.32 [0.21, 0.43] | 0.08 [0.02, 0.14] | 0.11 [0.04, 0.19] |
| 2B$_{_treatment}$ | 0.53 [0.44, 0.63] | 0.49 [0.39, 0.59] | 0.37 [0.28, 0.46] | 0.35 [0.26, 0.16] |
| 2C$_{_control}$ | 0.36 [0.25, 0.47] | 0.35 [0.24, 0.46] | 0.11 [0.04, 0.18] | 0.09 [0.03, 0.16] |
| 2C$_{_treatment}$ | 0.77 [0.72, 0.83] | 0.76 [0.70, 0.82] | 0.74 [0.68, 0.79] | 0.72 [0.66, 0.78] |

**Table 3.5** Out-of-sample $r^2$ comparisons between the original and alternative Combined Models

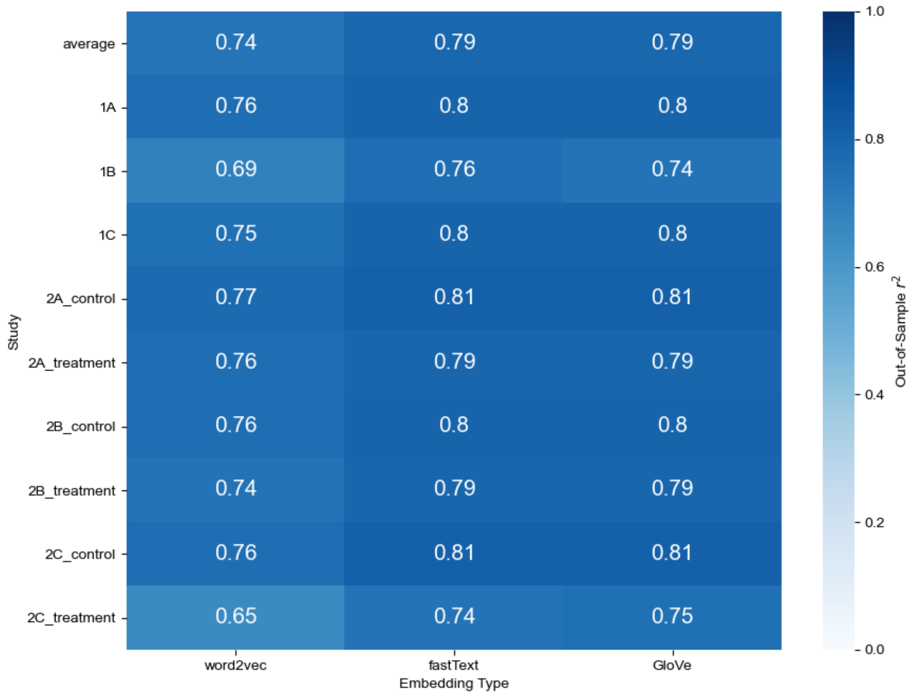
| Study | Combined Model (per 100g) | Combined Model (per 100 kcal) | Combined Model with micronutrients (per 100g) | Combined Model with micronutrients (per 100 kcal) |
|---|---|---|---|---|
| 1A | 0.77 [0.72, 0.82] | 0.76 [0.71, 0.81] | 0.66 [0.58, 0.74] | 0.74 [0.67, 0.80] |
| 1B | 0.75 [0.70, 0.80] | 0.72 [0.66, 0.77] | 0.59 [0.50, 0.68] | 0.67 [0.60, 0.75] |
| 1C | 0.76 [0.71, 0.82] | 0.75 [0.70, 0.80] | 0.68 [0.60, 0.75] | 0.73 [0.66, 0.79] |
| 2A$_{_control}$ | 0.77 [0.72, 0.82] | 0.77 [0.72, 0.82] | 0.66 [0.58, 0.74] | 0.73 [0.67, 0.80] |
| 2A$_{_treatment}$ | 0.81 [0.77, 0.85] | 0.79 [0.74, 0.83] | 0.73 [0.66, 0.80] | 0.77 [0.71, 0.82] |
| 2B$_{_control}$ | 0.77 [0.72, 0.82] | 0.76 [0.71, 0.81] | 0.66 [0.58, 0.74] | 0.73 [0.67, 0.80] |
| 2B$_{_treatment}$ | 0.83 [0.79, 0.86] | 0.81 [0.76, 0.85] | 0.75 [0.69, 0.81] | 0.79 [0.73, 0.84] |
| 2C$_{_control}$ | 0.77 [0.72, 0.82] | 0.77 [0.72, 0.82] | 0.65 [0.57, 0.73] | 0.74 [0.68, 0.80] |
| 2C$_{_treatment}$ | 0.91 [0.89, 0.93] | 0.90 [0.80, 0.92] | 0.88 [0.85, 0.91] | 0.89 [0.86, 0.92] |

### 3.6.4 Alternative Word Embeddings

In the main text, we fit the Vector Representation Model using pre-trained word2vec embeddings (Mikolov, et al., 2013) because these contained embeddings for multi-word food items. To show the robustness of using word embeddings in the Vector Representation Model, we also used other pre-trained word embeddings including fastText (Mikolov et al., 2018) and GloVe (Pennington et al., 2014), which both offer 300-dimensional vector representations of the food items. However, only 112 out of 172 food items have vector representation in fastText and 111 in GloVe. Unlike word2vec, which was trained on a large dataset of Google News articles, the GloVe model was trained on the Common Crawl 840B corpus and fastText on the Common Crawl 600B corpus. These embeddings also differ in terms of training algorithms. The pre-trained word2vec that we used was trained by a combination of the continuous bag-of-words (CBOW) method (which predicts words from their neighbors) and the skip-gram method (which predicts neighboring words of a given word). GloVe was trained in such a way that the dot product between two word vectors approximates the logarithm of the two words' probability of co-occurrence. FastText combines the CBOW method and position-dependent weighting to learn n-gram embeddings, which are summed to create word-level embeddings. Figure 3.8 shows out-of-sample $r^2$ of the Vector Representation Model (using ridge regression with $\lambda = 1$) with different semantic vector representations from different word embeddings for every study. As can be seen, the performance of the Vector Representation Model with fastText or GloVe is comparable and even higher than word2vec. We suspect the higher performance with fastText and GloVe is due to the fact that the training corpus (Common Crawl) underpinning these two pre-trained models contains a more diverse set of language data than that of word2vec (Google News).

**Figure 3.8**

Out of Sample Accuracy from Alternative Word Embedding Sources




*Note.* Out-of-sample $r^2$ of the Vector Representation Models (VRM) with different semantic vector representations from alternative word embeddings. The same set of 172 food items were used in VRM with word2vec. However, only 112 were used in the Vector Representation Model with fastText and 111 in GloVe due to the limited vocabulary of these embeddings.

### 3.6.5 Secondary Vector Representation Models

In order to address the issue of having a high number of potentially highly correlated predictors in the Vector Representation Model (and Combined Model), we used a ridge regression. To check whether our results are influenced by the choice of method, we tested other regression techniques including lasso, support vector, and k-nearest neighbors regression. Here we provide brief details about these techniques:

Both ridge and lasso learn a linear mapping from semantic vector representations ($x_i j$) to healthiness ratings ($y_i$) while penalizing the magnitudes of coefficients ($\beta_j$) to avoid potential collinearities between dimensions ($j$). However, they differ in the loss function used for estimating coefficients. Ridge penalizes the L2-norm of the coefficients (Eq. 1), whereas lasso penalizes the L1-norm of the coefficients (Eq. 2). Penalization parameter, $\lambda$, controls the strength of the penalty.

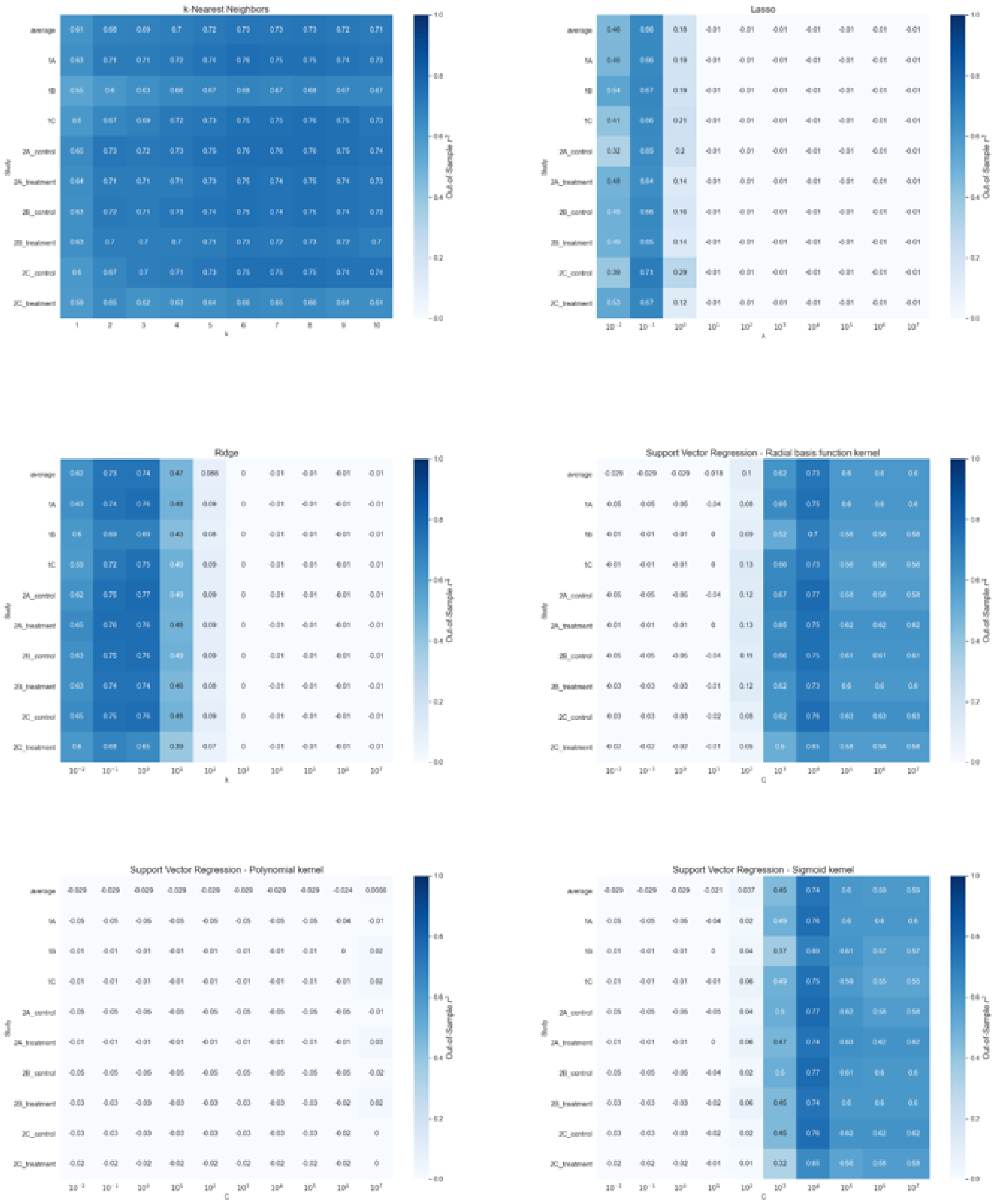
$$\Sigma_i(y_i - \beta_0 - \Sigma_j x_{ij}\beta_j)^2 +^{\lambda} \Sigma_j \beta_j^2 \tag{1}$$

$$\Sigma_i(y_i - \beta_0 - \Sigma_j x_{ij}\beta_j)^2 +^{\lambda} \Sigma_j |\beta_j| \tag{2}$$

Support vector regression uses a "kernel trick" to learn a nonlinear mapping from semantic vector representations to healthiness ratings with the penalization parameter, $c$, which works similarly to $\lambda$ in the lasso and ridge techniques. We considered three common kernel functions – radial basis function kernel (SVR-RBF), polynomial kernel (SVR-Polynomial), and sigmoidal kernel (SVR-Sigmoid). K-nearest neighbor regression predicts the healthiness rating of a food item by computing the average rating of the k nearest food items in the semantic vector space. The optimal penalization parameters $\lambda$ (in lasso and ridge) and $c$ (in SVRs) and the optimal number of neighbors $k$ are chosen through leave-one-out cross-validation. All analyzes were implemented in the Python scikit-Learn machine learning library (Pedregosa et al., 2011). For simplicity, keeping other tuning parameters in this library as default, we tested penalization parameters, $\lambda$ (in lasso and ridge) and $c$ (in SVRs), in the following set: $10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7$, and number of neighbors, k, in the following set: 1,2,3,...10. Figure 3.9 shows out-of-sample $r^2$ of different regression techniques with different tuning parameters for every study.

**Figure 3.9**

Out of Sample Accuracy of Secondary Vector Representation Models




*Note.* Out-of-sample $r^2$ of secondary Vector Representation Models with different penalization parameters.

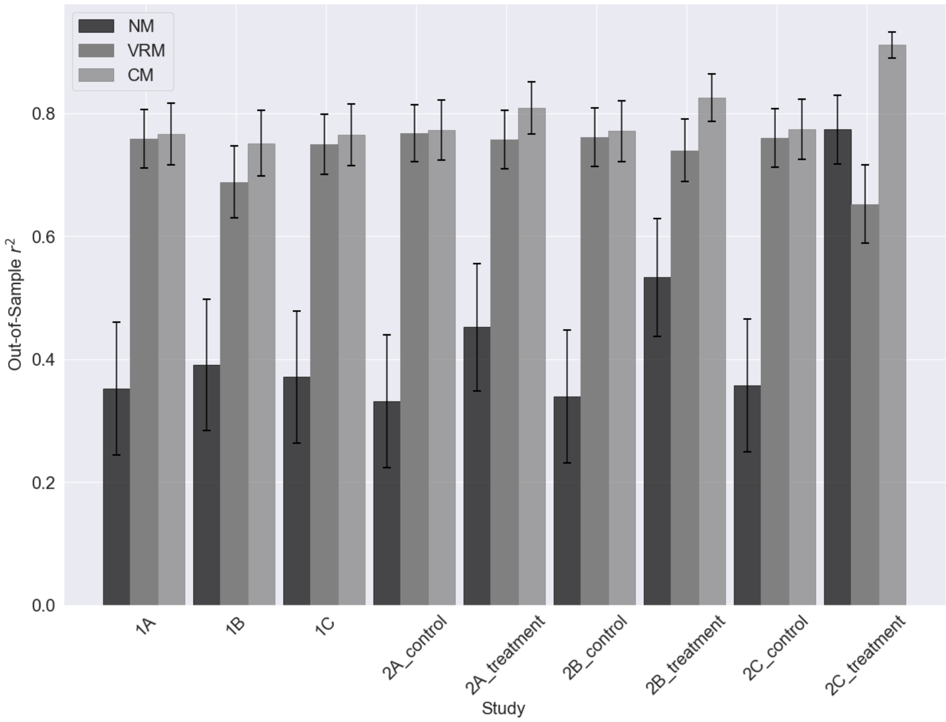
### 3.6.6 Test for Model Comparison

Figure 3.10 shows the out-of-sample $r^2$ with the 95% CI for the Nutrient Model, the Vector Representation Model, and the Combined Model trained on average healthiness ratings. Table 3.6 reports the Pearson correlation between the observed average healthiness ratings and the leave-one-out predictions, the out-of-sample $r^2$, and the mean squared errors of the Nutrient Model, Vector Representation Model, and the Combined Model. We can see clearly that both the Vector Representation Model and the Combined Model significantly outperformed the Nutrient Model except in the treatment condition of Study 2C where participants were given the most amount of nutrient information. We also ran separate paired sample t-tests to compare the squared errors from different models for each study. Across all studies, the mean squared errors from the Vector Representation Model and the Combined Model were significantly lower than those from the Nutrient Model (all $p < 0.01$). These results hold for the extended versions of the Nutrient Model and the Combined Model (detailed statistics are available upon request), suggesting that both the Vector Representation Model and the Combined Model outperformed the Nutrient Model.

Overall, the mean squared errors from the Vector Representation Model were not significantly different from those from the Combined Model (and its extended version, detailed statistics are available upon request), suggesting that adding nutrient information (on either macro- or micro-level) to the word vectors did not significantly improve the predictive accuracy of the Vector Representation Model.

**Figure 3.10**

Group Level Modeling: Model Comparisons




*Note.* Out-of-sample $r^2$ of the Nutrient Model (NM), Vector Representation Model (VRM), and Combined Model (CM). Error bars represent the 95% CIs.

**Table 3.6** Pearson correlation $r$, out-of-sample $r^2$ and mean squared errors (MSE) for the Nutrient Model, Vector Representation Model, and Combined Model

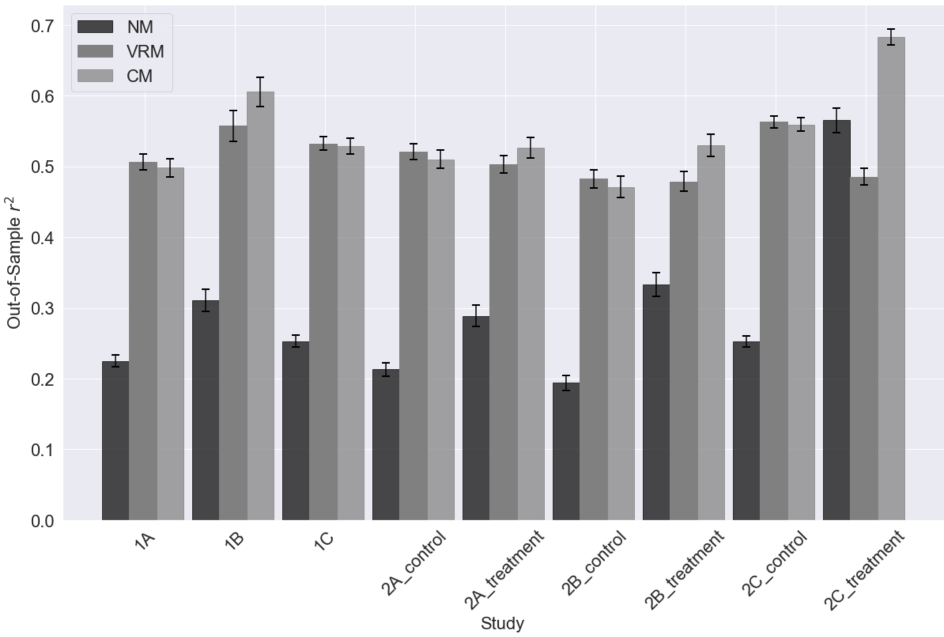
| Study | Nutrient Model | | | Vector Representation Model | | | Combined Model | | |
|---|---|---|---|---|---|---|---|---|---|
| | $r$ | $r^2$ | MSE | $r$ | $r^2$ | MSE | $r$ | $r^2$ | MSE |
| 1A | 0.60 | 0.35 | 1295.48 | 0.88 | 0.76 | 482.76 | 0.88 | 0.77 | 467.30 |
| 1B | 0.63 | 0.39 | 1471.25 | 0.84 | 0.69 | 753.24 | 0.87 | 0.75 | 600.55 |
| 1C | 0.61 | 0.37 | 1207.45 | 0.88 | 0.75 | 480.31 | 0.88 | 0.76 | 451.33 |
| $2A_{\_control}$ | 0.58 | 0.33 | 1353.49 | 0.89 | 0.77 | 471.76 | 0.88 | 0.77 | 460.46 |
| $2A_{\_treatment}$ | 0.67 | 0.45 | 1002.48 | 0.88 | 0.76 | 443.73 | 0.90 | 0.81 | 349.52 |
| $2B_{\_control}$ | 0.59 | 0.34 | 1183.60 | 0.88 | 0.76 | 427.76 | 0.88 | 0.77 | 410.59 |
| $2B_{\_treatment}$ | 0.73 | 0.53 | 828.56 | 0.87 | 0.74 | 461.70 | 0.91 | 0.83 | 309.75 |
| $2C_{\_control}$ | 0.60 | 0.36 | 1663.91 | 0.88 | 0.76 | 622.98 | 0.88 | 0.77 | 586.05 |
| $2C_{\_treatment}$ | 0.88 | 0.77 | 520.13 | 0.82 | 0.65 | 800.98 | 0.95 | 0.91 | 204.39 |

### 3.6.7 Individual-level Modeling

In the main text, we evaluated a Nutrient Model, Vector Representation Model, and Combined Model on group-level data, meaning that we averaged food healthiness ratings across all participants within each condition. However, since averaging ratings removes noise and variability, it is necessary to check if our results still hold on the individual level. To model individuals with the three types of models, we performed leave-one-out cross-validation on each individual's set of judgments. This yielded an out-of-sample prediction for each rating a participant gave. We then calculated the coefficient of determination ($r^2$) between these predictions and an individual's actual ratings and averaged these calculations across participants within studies. Figure 3.11 visualizes the out-of-sample $r^2$ for the three types of models for each study. The results on the group level are largely replicated on the individual level. First, the Vector Representation Model performs better than the Nutrient Model across all studies and conditions, except in the treatment condition of Study 2C. Second, the performance of the Nutrient Model increases as more nutritional information is provided in the treatment conditions of Studies 2A-2C. Third, the performances of the Vector Representation Model and Combined Model are indistinguishable in the general public sample of Study 1A and the control condition of Studies 2A-2C. Lastly, the Combined Model performs better than the Vector Representation Model in the expert sample of Study 1B and the treatment conditions of Studies 2A-2C.

**Figure 3.11**

Individual Level Modeling: Model Comparisons




*Note.* Average individual level out-of-sample $r^2$ of the Nutrient Model (NM), Vector Representation Model (VRM), and Combined Model (CM). Error bars represent standard errors.

### 3.6.8 Food Category Modeling

As a follow-up analysis, we applied our modeling to distinct categories of foods (e.g. baked products, beef products, dairy, and egg products). Overall, our results closely resemble our key findings reported in the main text, as can be seen in Table 3.7. The Vector Representation Model performs consistently high, with $r^2$ ranging from 0.76 (95% CI [0.72, 0.81]) to 0.83 (95% CI [0.79, 0.86]). By comparison, the predictive accuracy of the Nutrient Model was much lower but improved with the inclusion of more front-of-pack information across the studies ($r^2$ ranging from 0.27, 95% CI [0.16, 0.37] to 0.79, 95% CI [0.74, 0.84]). The Nutrient Model performed better than the Vector Representation Model when participants were shown traffic light labeling (Study 2C). However, the Combined Model achieves an even higher predictive ability than any individual model, with $r^2$ reaching 0.95 (95% CI [0.94, 0.96]) for the most informative (traffic light) labeling system.

We can also note the categories of foods that each of the models performs the best and worst in. There were no substantial differences between the predicted ratings and actual ratings for the Vector Representation Model in all the conditions except when judgments were made in the presence of traffic light labeling. In this condition, the biggest discrepancy between the model's judgment predictions and actual participant judgments was for the "Fats and Oils" category. Irrespective, both model predictions and observed ratings were always less than zero, meaning that the model accurately predicted that participants would perceive foods in this category as unhealthy. The Nutrient Model had the biggest discrepancy for the food categories of "Beverages", "Fats and Oils", and "Nuts and Seed Products", which was apparent in multiple conditions. As an example, for "Beverages", the Nutrient Model predicted healthier ratings than the actual ratings provided by participants in every case. This gap was still present when participants made judgments using traffic light labeling, despite the high overall predictive ability of the Nutrient Model in this condition. A possible explanation is that beverages like "cola" and "beer" do have a low negative nutrient content (e.g. fat, saturates, salt), hence why the model predicted a relatively high healthiness (above a rating of zero). However, in every condition except traffic light labeling, participants rated the "Beverages" category as unhealthy implying a distinct disparity between nutrient content and judgment formation. On the other hand, our Vector Representation Model always had a very high predictive ability for the category of "Beverages" demonstrating, particularly in this instance, that associations capture important attributes underlying healthiness judgments.

**Table 3.7** Out-of-sample $r^2$ comparisons between the Vector Representation Model, Nutrient Model and Combined Model for food category modeling, numbers in the brackets represent 95% CIs

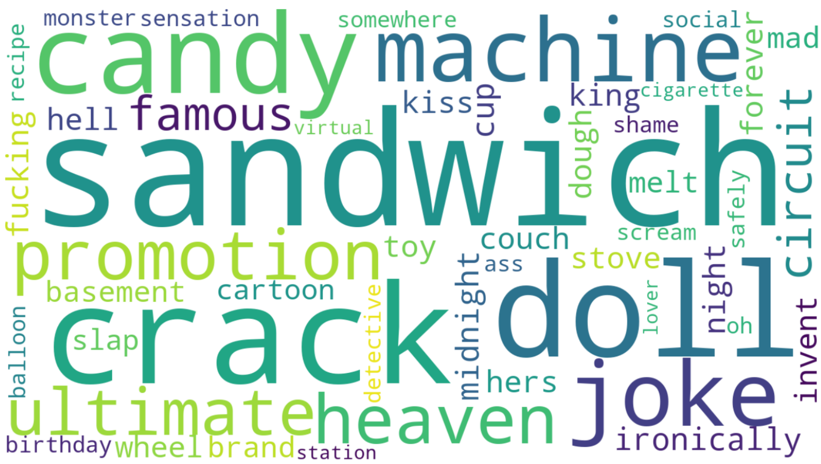
| Study | Vector Representation Model | Nutrient Model | Combined Model |
|---|---|---|---|
| 1A | 0.83 [0.79, 0.86] | 0.31 [0.20, 0.41] | 0.84 [0.81, 0.88] |
| 1B | 0.78 [0.74, 0.82] | 0.34 [0.25, 0.43] | 0.82 [0.79, 0.86] |
| 1C | 0.80 [0.76, 0.84] | 0.28 [0.18, 0.39] | 0.81 [0.77, 0.85] |
| $2A_{\_control}$ | 0.80 [0.76, 0.84] | 0.28 [0.17, 0.38] | 0.82 [0.78, 0.86] |
| $2A_{\_treatment}$ | 0.79 [0.75, 0.84] | 0.44 [0.35, 0.53] | 0.86 [0.83, 0.89] |
| $2B_{\_control}$ | 0.82 [0.79, 0.86] | 0.27 [0.16, 0.37] | 0.83 [0.79, 0.86] |
| $2B_{\_treatment}$ | 0.76 [0.72, 0.81] | 0.55 [0.46, 0.64] | 0.88 [0.86, 0.91] |
| $2C_{\_control}$ | 0.82 [0.78, 0.86] | 0.32 [0.21, 0.43] | 0.84 [0.80, 0.87] |
| $2C_{\_treatment}$ | 0.68 [0.62, 0.74] | 0.79 [0.74, 0.84] | 0.95 [0.94, 0.96] |

### 3.6.9 Word Cloud for Unhealthy Associations

It is possible to identify regions in vector space associated with unhealthiness in the same manner that we uncovered words associated with healthiness in the main manuscript (our Figure 3.4). As seen in Figure 3.12, the words with the lowest healthiness ratings (using a model trained on Study 1A participant ratings), have connotations that are more ambiguous. Despite this, there are recognizable terms related to less healthy categories of foods (e.g. “sandwich”, “melt”, “candy”, “alcohol”), and also words that seem to correspond to aspects of the context in which foods can be consumed (e.g.“midnight”, “couch”, “birthday”, “social”). While less intuitive, these associations that allude to the emotional and social benefits of unhealthy foods may explain why consumers choose them over options that are more nutritious.

**Figure 3.12**

Unhealthiest "Other" Words based on the Vector Representation Model




*Note.* Words with the lowest predicted healthiness ratings based on the Vector Representation Model estimated for the lay people in Study 1A.

### 3.6.10 Test of Model Generalizability

In order to establish whether our Vector Representation Model can provide true out-of-sample predictions about healthiness ratings, we collected participant ratings for an additional (not original 172) 60 foods (e.g. blackberries, custard, and spinach). To obtain this set of new stimuli, we generated model predictions (trained on Study 1A aggregate ratings) for all foods from the Food and Nutrient Database for Dietary Studies (U.S. Department of Agriculture, 2019). We then removed foods included in the original stimuli and uncommon or ambiguous foods. Following this, we selected 20 foods from the lowest quartile, 20 from around the median and, 20 from the highest quartile of predicted healthiness ratings, resulting in our final list of 60 foods. One hundred and one participants were recruited from Prolific Academic, with four participants removed using the same exclusion as the previous studies. There were 99 participants in our final participant sample (aged 19-61 years, $M_{age}$ = 26.71 years, SD = 8.98, 39.2% females, and 87.6% had no dietary restrictions). The design of this study was identical to Study 1A. The only difference was that participants rated the healthiness of food names that had not previously been used as stimuli. A paired-sample t-test was run to compare model predictions (trained on Study 1A data) with human ratings on both the aggregate and individual levels. On the aggregate level, the model predictions were not significantly different from the average participant ratings ($t(59)$ = -0.34, $p$ = 0.735). On the individual level, ratings from 53 out of 99 (53.54%) participants were not significantly different from the model's predictions. Based on these results, we conclude that the model can be used to predict subjective healthiness judgments for new foods, even in the absence of any nutritional information in the model specification.

# Chapter 4

# Discussion and Conclusions

How would you judge the healthiness of the food currently sitting in your fridge? It is likely that factors such as intended cooking method and accompanying foods would contribute to your evaluations, alongside presumed or visible nutrient information. These, and other, contextual factors are often overlooked in existing studies on healthiness perception accuracy, largely due to an inability of knowing which, and how many, of these attributes might be important. Two key recent advancements in computational linguistics (availability of large corpora and accessibility of vector space models) mean that psychology researchers have now discovered the value of computational approaches for understanding naturalistic judgments and choices (Bhatia, 2019; Richie and Bhatia, 2020). Inspired by such previous work, in this thesis, we use co-occurrences of different food names with other words in large-scale language data to uncover the most relevant attributes and associations for healthy (and unhealthy) food judgments (Bhatia, 2016, 2017; Bhatia et al., 2021a; Richie and Bhatia, 2020). Ultimately, we demonstrate how this data-driven methodology can progress our understanding of food healthiness perceptions.

## 4.1 Summary of Research Findings

In Chapter 2, we show how computational models trained on online news articles can uncover representations of superfoods. We chose to analyze superfoods specifically because it is a term that, despite being banned from appearing on food packaging since 2007 (EU Regulation No. 1924/2006), continues to be widely mentioned in relation to specific foods (Liu et al., 2021). Moreover, previous work suggests learned associations with superfood ingredients in food and beverage products cause health halo effects (Brownbill et al., 2018; Reisman, 2020), influencing consumer preferences (Meyerding et al., 2018). Yet, little is known about the nature of superfood representations. Therefore, in this chapter we used natural language processing methods to extract the words, concepts, and topics most likely to appear in "superfood" articles, with comparisons made against news articles about the same foods in non-superfood contexts. We also informally compare these frequent word co-occurrences with participant responses. Using three, independent, bottom-up text analysis techniques, we found that discourse surrounding superfoods in news media emphasizes and perpetuates health benefits, physical appearance, and disease prevention (despite weak scientific evidence in support of these claims). These associations, which emerge when using all three methodologies, showcase how pervasive the unsubstantiated associations with superfoods are.

In Chapter 3, we demonstrate how pre-trained computational models can be applied to predict food healthiness judgments. Across several experiments involving lay participants and trained dietitians, we find that our models are capable of making highly accurate out-of-sample predictions of food healthiness ratings for a wide range of foods. Additionally, we extend previous similar research by also investigating the susceptibility of judgments to observable information often found on food packaging, in this case, nutrition labeling. As a result, we provide a tool that has the potential to assist in the evaluation of behavioral interventions intended to shift individual food healthiness perceptions in the right direction. A further advantage is that our models are highly general, and so can be used to predict judgments (and intervention effects) for nearly any food item. Indeed, we share our model's predictions for hundreds of novel food items as we believe it is a useful resource for future research. We also shed light on the black box underlying the generated predictions of our machine-learning approach. This conceptual mapping is achieved by applying the same model to predict the healthiness ratings of common words from an established semantic dictionary. Here, we uncover an entwined relationship between high healthiness perceptions and naturalness related words. Moreover, for the first time, we find a role of social and emotional contextual factors in underpinning perceptions of foods as low in healthiness.

## 4.2 Theoretical Implications

Theoretically, while we continue to make clear that vector-based representations are not themselves a model of psychological processes (Bhatia et al., 2021a; Richie and Bhatia, 2020), such a mechanism of associative judgment is cognitively plausible. This is a strength of the approach, as it acts as a continuation of existing theoretical research (Bhatia, 2019). Indeed, cognitive scientists have effectively used vector space models to predict both low-level behavioral phenomena and high-level judgments in other domains (Bhatia, 2019; Bhatia et al., 2021b). The consistently high accuracy of our vector representation model in predicting out-of-sample judgments of independent groups of participants, can therefore offer a less costly method of obtaining insights into latent attributes (Bhatia et al., 2021a). While a perfect representation of people's knowledge about foods is unlikely, the gradual exchange of predictive power from the vector representation model to the nutrient model suggests vector representations capture some truth. Moreover, this underlying human-like mechanism, along with a high predictive power of our integrated model (Combined Model), implies the possibility of an interaction between the two (currently perceived separate) models of thought (Bhatia, 2017; Kahneman, 2003; Kahneman and Frederick, 2002; Morewedge and Kahneman, 2010). This reinforces the idea of a single system of learning, storing, and retrieving knowledge, as suggested by Bhatia (2016). Thus, at the very least, our findings bring forth an exciting avenue for further research into the interplay between knowledge representations and observable information.

## 4.3 Practical Implications

### 4.3.1 Generalizability of our Predictive Model

The most significant practical implication of our research is the development of a computational tool that can determine the influence of package features in shifting healthiness perceptions of individual foods. Our primary achievement was in creating baseline predictions for hundreds of commonly consumed foods trained on ratings using food names only (see https://osf.io/jys6u).

As demonstrated through Studies 2A, 2B, and 2C in Chapter 3, judgments of healthiness can also be predicted when participants are presented with energy and/or nutrition labeling (in various formats) alongside each food name. Our findings that traffic light labeling is the most effective, out of the three types, in changing judgments is in line with previous studies (An et al., 2021; Cecchini and Warin, 2016).

Indeed, policy-makers have been advised to avoid non-interpretive label formats (like our monochrome version), despite being preferred by industry, because many consumers struggle to understand and use them (Jones et al., 2019). Nonetheless, notably, we can now measure the extent of influence at a food or food category level, and specifically, whether foods change from being considered as healthy at baseline to unhealthy after intervention (or vice versa).

On a side note, one may have noticed that the reference intake guides were absent from our monochrome and traffic light labeling designs. We chose to implement design features singly to ascertain the independent influence of each feature. This would allow policy-makers and certified nutritional experts to deliberate over the fewest features necessary to nudge the public effectively. Indeed, we encourage researchers to test different combinations of features from existing nutrient labeling formats, including those from other strategies (e.g. health stars, nutritional warnings, stop signs) used around the world (Jones et al., 2019).

The ideal scenario would be to ascertain how judgments culminate when lay decision-makers are presented with all food packaging features, including marketing tactics, found in real-world environments. This would be the most effective approach for an evidenced-based method to determining a single, universal, front-of-pack nutrient labeling strategy. Furthermore, it would offer valuable insights into how consumers trade-off conflicting government-provided and marketing-provided information.

### 4.3.2 Potential for an Evaluative Objective Healthiness Tool?

It is also important to highlight that the aforementioned approach is not an evaluative tool itself. Nutritional experts would be needed to discuss and ascertain the accuracy of (predicted) healthiness ratings. Instead, the novelty is in being able to generate accurate predictions for hundreds or thousands of foods after being trained on a small sample of human ratings for approximately 200 foods. Moreover, a strength is that these predictions can be made without any objective model or nutritional composition data.

With that said, the high predictive power of the Combined Model shows that it is also possible to incorporate nutrient composition data (e.g. from the very recent Food Compass policy tool (Mozaffarian et al., 2021) with vector space models, or run a Nutrient Model alongside, if a key-nutrient perspective is required.

Relatedly, one might anticipate the question of whether such a computational approach can be used to create an objective, evaluative, individual food healthiness model? As it stands, computational models are only as good as the data that they are trained on. Unfortunately, there are statistical regularities in natural language corpora that mean historic biases and prejudices are mirrored in word embeddings (see Arseniev-Koehler and Foster (2020); Caliskan et al. (2017); Charlesworth and Banaji (2021)). Thus, a machine-learning approach is unable to provide us with a pure and neutral output (Caliskan et al., 2017). However, in contrast, for approaches like ours that look to mimic human behavior and judgments, these are the exact systematic biases that are of interest and value to our research.

## 4.4 Suggestions for Future Research

### 4.4.1 Use of Color

An advantage of using data-driven bottom up models is that the uncovered associations and biases can provide direction to future research on healthiness perceptions. When informally comparing the superfood and organic associations from Chapter 2 with the latent associations reflected in the word clouds from Chapter 3, we notice thought-provoking similarities and differences. As expected, reference to nutrient content was reflected in all analyzes. The heuristic of color, particularly green, was also found in all three outputs. Interestingly, the construct of naturalness appears to have a prominent association with healthiness for organic representation and in judgment prediction, however it had a smaller role in superfood representation than expected. One stream of thought is that use of the color green may have a larger role in healthiness perceptions than currently thought. For instance, both "organic" and "fairtrade" labels have been found to increase perceptions of healthiness, despite being unconnected to nutrient content (Berry and Romero, 2021; Schuldt et al., 2012). As both certified labels contain green in their design, it is conceivable that this use of color may contribute to their perceived association with healthiness (Carlsson et al., 2021; Schuldt, 2013).

### 4.4.2 Importance of Contextual Factors

Another interesting finding is in reference to the associations underpinning the model predictions of foods rated as low in healthiness (see Supplementary Materials in Chapter 3). First, it suggests that associations that influence perceptions of a food as unhealthy are not necessarily the opposite of those for healthiness perceptions. In fact, many of these terms appear to be associated with transgressional or deprecated behaviors (e.g. "fucking", "cigarette", "crack", "slap"). Second, a reference to both "heaven" and "hell" implies that unhealthy foods evoke stronger religious associations than healthy foods. Considering the frequent language about certain foods being denounced as sinful or immoral in the past (e.g. red meat, salt, spices), one may wonder if historical representations about foods persist in language (Oakes, 2004).

### 4.4.3 The Evolution of Healthiness Representations over Time

Other computational approaches are emerging that may allow future researchers to capture changes in food healthiness representations over time. As it stands, there are several corpora that allow us to trace the historical meanings of words over a sufficiently long period of time (Li et al., 2020). Among the most far-back reaching ones are the Google Books Ngram Corpus (books published from 1600-2009) (Lin et al., 2012); followed by "Find My Past" data from the British Library's "British Newspaper Project" (newspaper articles from 1710-1953); as well as the Corpus of Historical American English (COHA), which consists of 400 million words produced from 1810s to 2009 (Davies et al., 2012).

Whilst these corpora have shown considerable success for uncovering reliable historical word embeddings for well-established terms such as wellbeing and risk (Hills et al., 2019; Li et al., 2020), historical data for niche food healthiness terms such as "superfood" are likely to be too sparse to reveal meaningful patterns. For instance, "Find My Past" has only 523 articles mentioning "superfood". To overcome this issue, we would recommend capturing evolving representations of specific food names (e.g., "banana", "coffee", and "avocado"), similarly to our approach in

Chapter 2. This approach would be more likely to provide the thousands of relevant articles needed across periods of months or years (Bhatia et al., 2021a). However, it should be acknowledged that the Google Books corpus only contains a digitized historical record of approximately 6% of books published, so some food representations may still occur too infrequently.

Irrespective, computational scientists have recently introduced dynamic word embedding models (Yao et al., 2018), which are an extension to the word embedding models (e.g. word2vec and GloVE) discussed in this thesis. Essentially, dynamic word embeddings can take corpora from different time periods and learn corresponding vector spaces for each time slice (Richie and Bhatia, 2020). Hence, these more recent models show promise for allowing researchers to infer future social trends and changing word semantic associations over time (Yao et al., 2018). Therefore, with the increase in news articles since the start of the digital age, there will be new opportunities to train embedding models in real-time (Bhatia et al., 2021a). As a result, such an approach could be used in the creation of a policy tool to recognize, flag and dispel myths arising from harmful wellness and diet trends (e.g. raw unpasteurized milk, appetite-suppressant lollies, and detox teas), which have unfounded associations with healthiness.

### 4.4.4 Unintended Consequences of Current Healthiness Discourse for Unhealthy Eating Behaviors

Many of the associations that underlie both the high and low healthiness perception predictions could be seen as entering into the territory of disordered eating behavior discourse. It is the emphasis on naturalness for healthy foods and transgression-implying words for unhealthy foods that suggest a connection with "clean eating" beliefs in particular. To clarify, clean eating is a diet where followers eat only non-processed, organic, plant-based, raw, "real", local or home-cooked foods (Ambwani et al., 2019). This often also includes eliminating certain foods or food groups (gluten, grains, dairy) from their diet without justification, and is not endorsed by certified nutritional experts (Ambwani et al., 2019).

In fact, one noteworthy observation when informally comparing associations generated by the predictive model trained on registered dietitians ratings (see Figure A.1 in the Appendix) with the models trained on lay decision-maker ratings, is a difference in intensity of the words associated with unhealthiness. For instance, words such as “slap”, “fucking”, “scream” are not present in the word cloud for Study 1B, although references are still made to social and environmental context of unhealthy food consumption (e.g., “midnight”, “drinking”, “promotion”. On the other hand, we do observe the same nature-related associations with foods rated as high in healthiness. Nevertheless, one could speculate that while associations for healthy and unhealthy foods are relatively similar, the difference between healthy and problematic perceptions is in the strength of these associations.

This would line up with research suggesting that people fail to recognize the complexity of healthiness in reference to individual foods, preferring to classify them as either “good” or “bad” (Freeland-Graves and Nitzke, 2013; Green, 2015; Julia et al., 2021). As a consequence, there is a general tendency to underestimate the calories of perceived healthy foods and vice versa (Bui et al., 2017; Larkin and Martin, 2016). Indeed, this oversimplification ignores how some nutrient-dense foods can be caloric (Nicklas et al., 2014), and how consuming less energy dense foods only can have consequences for the intake of some essential nutrients (Nicklas et al., 2008). Relatedly, purported superfoods often do have high nutritional content, but the problem is claims about disease prevention, anti-aging, and weight loss benefits become far-fetched (Curll et al., 2016; MacGregor et al., 2021).

However, further scientific studies, and input from disordered eating specialists, would be required to make any further comment on this topic. Nonetheless, our research highlights a need for interdisciplinary collaboration to, at least, ensure the complexities of healthful eating behaviors at a population-level are considered when evaluating the effectiveness of policy tools.

### 4.4.5 Healthiness Predictions of More Words

It may also be of interest to know that the conceptual mapping from our predictive healthiness model can be applied to a nearly unlimited set of words (Bhatia et al., 2021a). The aforementioned word clouds were trained using the COCA (Corpus of Contemporary American English) dictionary. However, if desired, the model can predict the healthiness of any list of words in the same manner as out-of-sample food predictions. This means that researchers can obtain a list of healthiness ratings from marketing terminology, words with known valence and arousal ratings (Warriner et al., 2013) to known sensorimotor norms (Lynott et al., 2019). Consequently, our approach allows for detailed examination of all commonly used words and their relationship to either high or low healthiness perceptions.

We would encourage this approach over comparing vector similarities between words of interest (e.g. interpreting the cosine similarity difference between the vector for "healthy" and the vector for a word like "natural" or a food like "avocado"), such as Grand et al. (2018). This latter method has received criticism, largely because it wrongly implies cosine similarity, and vector representations, are a model of the cognitive process underlying human behaviors (Griffiths et al., 2007; Jones et al., 2018; Richie and Bhatia, 2020). Whereas, a strength of our computational approach to obtaining associative words is that it avoids that assumption (see Richie and Bhatia (2020) for detail).

### 4.4.6 Individual and Group Level Differences

Future researchers may also want to further explore individual-level idiosyncrasies and heterogeneities in food healthiness representations across different population groups. While we demonstrate that the pre-trained word2vec model performs well for participant-level modeling, other predictive computational models may be better suited to answer healthiness perception research questions about differences between groups. For instance, Singh et al. (2020) highlights the potential of a more sophisticated machine learning technique, called a multilayer perceptron, that can project additional covariates (demographic variables, psychographic variables, EQE-Q (eating disorder questionnaire) results) into the hidden layers of the regression. These hidden layers then learn how various individual characteristics interact with judgments, and thus can provide insights into group-level variability (Singh et al., 2020). Many existing studies have found mixed results over whether gender and sex is a contributing factor to differences in food healthiness perceptions (Carels et al., 2007; Foroni et al., 2022; Oakes and Slotterback, 2001, 2002), thus this would be a possible avenue to investigate potential differences.

Of particular interest, however, is the influence of age. Perkovic et al. (2021) reported concerning preliminary findings of binary perceptions of foods in adolescents. Therefore, it would be fascinating to explore whether this difference in food healthiness perceptions of adolescents is also picked up using our computational approach, and whether the latent associations remain the same. Indeed, Chapman and Maclean (1993) found that, for adolescents, consumption of junk food was associated with independence, pleasure, and friends whereas healthy food was associated with parents and being at home. Therefore, it would be of interest to know if the high levels of unhealthy food consumption reported in teenagers is based on well-informed decisions, or knowledge errors and biases in perception. Crucially, this finding would be influential in the development of interventions aimed at curbing obesity in this age-group, as it is possible that highlighting the unhealthiness of certain foods only makes them more attractive. Subsequently, it would establish whether a more valuable approach would be to instead focus on food industry regulation and encourage more food and beverage product reformulations (Gressier et al., 2020).

Some biases of our chosen news corpora will likely impede the generalizations we can make to understanding universal food healthiness perceptions. For one, use of English-speaking corpora means that our models are more likely to capture western representations of foods (Bhatia et al., 2021a). For instance, one study found that the western superfood, amaranth, was seen traditionally as a basic food, used in the production of "Alegría", a Mexican candy (Rojas-Rivas et al., 2019). Similarly, acai berry, one of our top 25 superfoods reported by consumers, was traditionally a staple food for Amazonian peasants where it was eaten alongside fish and, cassava flour or grits (Parker et al., 2019). In fact, traditional beliefs were that acai berry caused lethargy, opposite to the western messaging associated with acai where it is advocated for increasing energy (Parker et al., 2019). Therefore, it would be interesting to assess whether our findings are replicated when using various non-English corpora, task instructions and food names.

Another factor is the representativeness of these corpora to different population demographics. It is possible that our focus on news articles is more likely to align with the views of people who read news articles. Studies have shown that adolescents are more likely to obtain information from social media channels, food industry brands, YouTube, and relatable influencers than from those with expert nutritional knowledge (Barklamb et al., 2020; Coates et al., 2020; Pilgrim and Bohnet-Joschko, 2019; Qutteina et al., 2019). These sources likely ignore the language nuances of certified nutritional advice that tends to be kept in regulated sources of information (e.g. "avoid for better health outcomes" vs the correct "eat less for better health outcomes") (Adamski et al., 2020; Ramachandran et al., 2018; Rousseau, 2015). Moreover, the majority of posts from celebrity social media accounts show (and arguably normalize) unhealthful eating behaviors (Turnwald et al., 2022). Therefore, future researchers should also consider the representativeness of the news corpora to target populations, particularly for individual or group level modeling.

### 4.4.7 Exploring Perceptions of Other Food Choice Motivators

While the focus of this thesis was on food healthiness perceptions, the approach can be directly transferred to study perceptions of other motivators of food choice. For instance, consumers are becoming increasingly concerned with the environmental impact of their diet (Delgado et al., 2021), Yet, little is known about consumers perceptions of relevant concepts, like food edibility, which can reduce food waste (Nicholes et al., 2019). Additionally, neither is it clear how accurate people's knowledge is about concepts like carbon footprint for different foods (Armstrong et al., 2020). It would be straightforward to replicate our analysis, simply by changing the question phrasing from "...rate the healthiness..." to "...rate the environmental impact..." for our 172 food stimuli. Subsequently, researchers can input the new participant ratings into the available python script to add predicted perception ratings of environmental factors for the hundreds of food names we made available.

However, for the instances where perceptions of objective, fixed, numerical data is desired (e.g. carbon footprint, greenhouse gas emissions, water footprint), a more suitable approach might be to use judgment error models (Zou and Bhatia, 2021). In a recent study where participants were asked to estimate the calorie content of foods and infant mortality rates in various countries, Zou and Bhatia (2021) found that, in both scenarios, errors were more likely to be a result of using incorrect knowledge, rather than a difficulty in applying knowledge (as ascertained from the predictive power of different computational models). Here, difference ratings between the participant ratings with the objective rating allowed insights into the associations that lead to over-estimations and under-estimations. Hence, the best computational approach would depend on the intricacies of the research question.

Along these lines, other computational approaches may be more appropriate for other research questions in the food domain. For instance, compositional distributed representations, such as BERT and ELMO (Devlin et al., 2018; Peters et al., 2018), can vectorize sentences rather than words. This allows them to account for sentence structure and incorporate more context into their embeddings. As a result, these methods have the added benefit of recognizing identical words with multiple meanings such as "orange" or "squash", which has been shown to improve performance (Richie and Bhatia, 2020). As such, these newer generation models have been found to be well suited for tasks where researchers might want participants to select a span of text from a passage (e.g. perceived important information from a recipe or meal description) to answer a question (Richie and Bhatia, 2020). Conversely, such an approach is unable to train representations of multi-word phrases such as “apple pie” (Bhatia et al., 2021a), which we felt more important for our particular research question. Irrespective, a strength of our research, in testing vector representations using multiple computational models, is that we provide further support for the robustness and relative similarity in predictive power across different computational models (Bhatia, 2017).

## 4.5 Conclusions

In conclusion, how do we help people make better, healthier food choices? Our response, and contribution to the literature, can be summed up in three key points. First, find an approach that can capture (or approximate) the rich knowledge that people hold about foods. Second, uncover the attributes that can explain, and predict, the variability in healthiness judgments of different foods. Third, create an evidence-based approach that can measure changes in healthiness perception caused by different food package features (for all commonly consumed foods). It is only at this point that we can assess the effectiveness of interventions aimed at improving healthy eating in the public, as well as monitor any unintended consequences of simplified messaging. Our approach highlights that lay decision-makers, and registered dietitians, consider non-nutrient attributes important to food healthiness. Indeed, foods are more than the sum of their nutrients (Forouhi et al., 2018). Therefore, measures of perception accuracy should equally consider the importance of mental, social, and physical health associations in people’s dietary patterns and eating behavior. This holistic approach to healthy eating is encompassed within the broad guidelines of “balance, variety, and everything in moderation” and food-based (rather than nutrient based) guidelines such as the EatWell Guide (Green, 2015). However, more needs to be done to empirically define these terms or provide reference points, as the interpretation of these terms varies greatly among consumers (Paquette, 2005; vanDellen et al., 2016). Nonetheless, exaggerated associations have emerged since the shift onto individual food healthiness. In truth, we should be wary when following nutrient labeling (a tool to assist healthy eating behaviors) obsessively, or rather perfectly, leads to the solution becoming part of the problem. Available packages such as “gensim” and more recently, “magnitude” make pre-trained embeddings for many word computational models accessible to researchers new to natural language processing (Patel et al., 2018; Řehůřek et al., 2011). This is particularly advantageous to psychologists and behavioral scientists who wish to explore and extend our research in this domain. Indeed, the research shown in this thesis is just a drop in the ocean of the potential that computational approaches have in transforming our understanding of food healthiness. Nonetheless, one thing we have established (using a variety of computational methods) is that linguistic co-occurrences in online news articles can remarkably reflect what people know about foods. This means that associations in media, created to sensationalize and simplify complex information, likely contribute to the strength of learned associations in lay decision-makers' minds. Ultimately, if only one thing could be remembered from this thesis it would be, language matters.

# Chapter 5

# Further Reflections

In this section, I will provide a reflective account of my experience conducting the aforementioned empirical chapters. Specifically, I will cover additional considerations, constraints, and lessons I have learned, with the goal of helping future academics in this field who may wish to replicate and/or extend this research.

## 5.1 Superfood Representation in News Media

### 5.1.1 Ambitions

The inspiration behind this paper was an online data visualization from the Information is Beautiful blog entitled “Snake Oil Superfoods” (McCandless et al., 2018). The visualization is a bubble plot of 81 unique superfood names, rated on the level of scientific evidence supporting a specific health benefit claim. The size of each bubble indicates the claim’s popularity, measured by Google hits. I was intrigued by the high popularity of certain foods associated with having a beneficial impact on health, despite there only being “slight evidence” or “no evidence” reported. For example, açaí berry is regarded as a food to help with weight loss but there is no definitive scientific evidence in support of this claim. The information presented in this visualization formed the premise of my ambitions to investigate how unsubstantiated associations with certain foods and health are formed.

### 5.1.2 Corpus Challenges and Limitations

Before choosing the approach that formed my final analysis in Chapter 2, my collaborators and I first considered other avenues for collecting superfood representations in the media. This included exploring whether the pre-trained word2vec model (used in Chapter 3) could capture the unsubstantiated associations found within the "Snake Oil Superfoods" visualization. Using the Vector Representation Model trained on participant ratings from Study 1A and the control conditions of Study 2A-2C, I found that the mean rating of healthiness for these foods was far lower than I expected. To explore this further, I checked the top 100 most similar vectors for each of the 41 superfoods using the pre-trained word2vec corpus in Python (Mikolov et al., 2013). I also looked into the similarity between the word embeddings for each superfood and the associated condition mentioned in the "Snake Oil Superfoods" visualization. These results implied that the pre-trained word embedding model was likely trained using a small window size, which is better at capturing functionally similar words than the domain or topic of each word. The low cosine similarity scores also suggested that the "superfood" term may not be adequately represented in the corpus. Subsequently, I concluded that the Google News corpora, trained on articles written before 2012, may not be the best corpus to use for analysis of niche topics such as superfoods. As a result, I pivoted towards finding another corpus, which contained articles written more recently, to undertake my exploratory analysis.

My main priority was to select a corpus that could capture how representations of superfoods might be formed at a population level. I chose the NOW Corpus (http://corpus.byu.edu/now/), as articles from magazines and newspapers have a broad reader base and are used to disseminate information about topics and trends of interest to the public. I also felt that the size of the corpus and its date range (approximately 10 billion words of data between January 2010 to February 2020) would provide me with access to a large sample of the most up-to-date articles for my analysis. This also raised the possibility of exploring changes in superfood representation over time, which was part of my original ambition. However, the final sample size of superfood articles was too small for this analysis.

At this stage, I thought the best approach would be to explore all articles mentioning the word "superfood". This would allow me to establish the largest sample size of potentially relevant articles in the corpus. Surprisingly, this produced just 1226 articles despite selecting only those from the United States of America or Great

Britain, and those containing the words "food" or "diet", as well as "superfood" at least once. When manually checking the articles, I also noticed several irrelevant articles had been included (e.g., detailing an accident that had occurred outside a superfood store). As a result, my collaborators and I discussed various machine-learning techniques that could help me obtain a greater number of relevant articles. We first applied a natural language processing technique known as Doc2Vec, which involved training a model on all the superfood articles in the NOW corpus, treating each document as a vector. This approach used outlier encoding to try and rebuild the same vectors for each document, with the goal of using only the most influential coefficients but minimizing the construction error. I decided not to take this approach as manual inspection showed that some relevant articles were incorrectly excluded. The second approach involved averaging pre-trained vectors for each word in the title or text of all food-related articles, weighing them by frequency. Next, using the first principal component of each article vector representation, we ran a similarity check with the vector representation for the word "superfood". However, it was agreed that this method would be biasing the sample and it would be better to use simpler methods of establishing the initial sample of superfood articles. Subsequently, I decided on selecting articles that mentioned at least one food name from the food name list as well as the word "superfood" in either the text or title. Although there was a trade-off between finding all superfood articles, I felt this was the best approach to reduce researcher bias yet improve the relevance of the articles included.

The main limitation of the NOW corpus was that it did not contain all relevant articles during the specified time frame as expected. This was discovered after carrying out a manual Google News search, using the same dates as the NOW corpus dataset, to investigate the cause behind the low sample size of relevant superfood articles. This investigation found many article titles that were missing from the NOW corpus, even when only filtered on articles containing the words "food" or "diet". It is unclear how many articles are missing in the NOW corpus, but this is a consideration that will have had a limiting factor on the representativeness of my findings.

Relatedly, I assumed that this corpus would provide a representative sample of articles about superfoods from a broad range of news outlets. In hindsight, I should have run a frequency analysis to visualize the distribution of news outlets represented in the corpus. The reason is that the news outlet plays a role in the language used and would increase the reproducibility of my findings in the future.

Indeed, Figure 5.1 shows the news outlets from which at least two relevant articles were published. Awareness of this limitation earlier would have permitted actions to rectify any imbalance of articles from certain news outlets by collecting articles from additional corpora, with less risk of this being seen as a response to unexpected findings. I would also recommend that future researchers consider whether it is possible to obtain information about the article's author as well as the type of outlet (e.g., broadcast, tabloid, magazine). This would allow for a more rounded and insightful account of the media sources used to conduct the analysis.

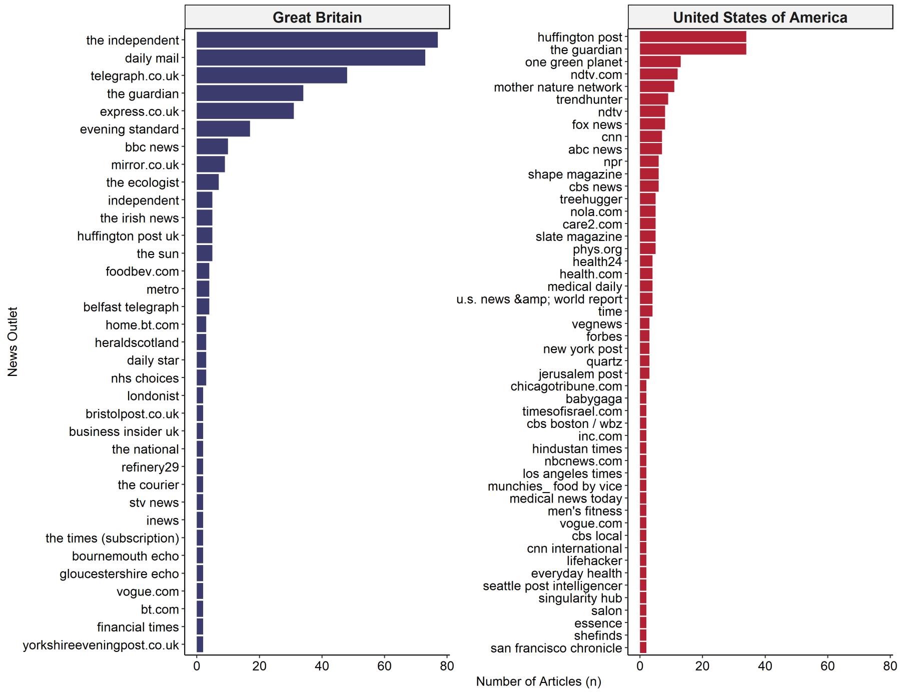



**Figure 5.1** The frequency of articles from different news outlets, separated by country of origin (Great Britain or the United States of America)

It was imperative to identify a comparison group of articles to make any claim that my findings were unique to superfood representation in news articles. As a result, I decided to focus on a subset of foods that are commonly perceived as superfoods and match articles about them written in different contexts. However, unfortunately, there is no definitive list of superfoods, with a wide range of articles written about different foods in recent years. I first considered using the list of 41 foods from the "Snake Oil Superfoods" visualization, but I was unsure of the true origins and accuracy of this data. Therefore, I ran a frequency analysis to find the most commonly mentioned foods across all superfood articles, and see how the words compared. I discovered that the most frequent foods did not align with the superfoods from the "Snake Oil Superfoods" dataset. Thus, as a sanity check, I decided to collect participant data to determine a list of the most recognized superfoods by the public. As the participant data produced a different list of superfoods, I concluded it was best to go forward using the list of 25 foods from the participant data collection. This was decided because, at the time, I felt the superfoods mentioned by the sample of participants were the most reliable out of the available options. Reflecting on this, the sample size used to collect the participant data was relatively small, and not based on any previous literature. If this research was to be replicated, one recommendation would be for researchers to follow my method but collect participant data on a larger sample to increase the accuracy and reliability of the foods identified.

My decision to include articles about the same foods in an "organic" context was to show that my approach was not just capturing healthiness representations in the broadest sense. I considered multiple terms that are related to a different, but important, aspect of food healthiness perceptions such as "natural", "vegan", and "raw". However, I felt that "organic" was the least ambiguous term that could demonstrate the subtlest of differences in the representation of the same foods written in a superfood context. There is a considerable overlap of 194 articles that mention both the "superfood" and "organic" terms. I did not exclude these articles from my analysis and it is possible that they appeared twice in the two separate contexts. Nonetheless, 4658 articles mentioned the term "organic" and not "superfood", whereas 679 articles mentioned "superfood" and not "organic". For future researchers, if they were able to obtain a larger sample size I would recommend removing the articles that overlapped to have a clearer distinction of foods from each group.

## 5.2 Predicting Food Healthiness Judgments

### 5.2.1 Ambitions vs Reality

The ambition for this paper was to uncover how healthiness judgments of individual foods are formed, and the extent to which attributes often found on food packaging influence healthiness perceptions. Once it was established that a vector space approach could accurately predict healthiness judgments when participants were presented with a food's name only, I wanted to investigate whether such judgments could be shifted. I aimed to incrementally increase the information presented to participants to eventually reflect naturalistic food packaging designs (including nutrient labeling, nutrient and health claims, food descriptors, branding, ingredients list, product color, and product shape). This would allow for a rigorous and evidence-based assessment to ascertain the influence of all food packaging attributes on healthiness perceptions.

The reality, however, is that many elements within each design component may shift the overall healthiness perception of a given food to varying degrees. For example, as demonstrated in the results section of Study 2A to 2C in Chapter 3, interpretative traffic light colors were most influential in aligning people's healthiness judgments with key nutrient content predictors. Hence, it is important to isolate the effect of individual elements in each of these aforementioned design components to identify simple changes to packaging that can help consumers make better-informed choices. Thus, one can imagine the complexity of assessing the influence of each design feature, both independently and in combination with other food packaging features. Nonetheless, it is only then that an overall picture of how attributes are traded off and healthiness perceptions shift for foods across all categories can be obtained.

Consequently, I chose to focus my research on the one aspect of food packaging that is designed to improve healthiness perceptions, nutrient labeling strategies. Indeed, there are many nutrient labeling strategies around the world, but there is limited research directly comparing their influence on people's healthiness perceptions for a variety of foods across different food categories. Therefore, I felt that focusing on this particular component would allow me to highlight the potential of using a vector-based approach to evaluate the effectiveness of different nutrient labeling designs. Moreover, I decided to share healthiness predictions for hundreds of food items using the Vector Representation models trained on participant ratings from each of the conditions in the main study. It is hoped that other researchers will extend this work by testing the influence of other food packaging features, as per my initial ambitions.

Nevertheless, it is worth mentioning that I have also collected data to investigate the influence of three non-nutrient labels (that are often associated with healthiness) in shifting healthiness perceptions. The following subsection will provide an account of this additional study, alongside my reflections from this body of research.

### 5.2.2 Additional Study

Studies 2A-2C demonstrate how people can, and do, incorporate nutritional information into their healthiness judgments when provided. The consistently high performance of the Vector Representation Model shows people still rely on associations with the food name. Typically, food companies also add non-nutrient-related front-of-pack labeling, which consumers sometimes mistakenly take as an indicator of the food product's healthfulness (Berry and Romero, 2021; Richetin et al., 2022). Therefore, I also ran this additional study to investigate how further associations created by such labels, either highlighting food origin ("organic") or ethical considerations in food production ("fairtrade"), might influence the performance of the computational models. Moreover, whilst the use of the term "superfood" is banned on food packaging, in Chapter 2 we found the term superfood to be highly associated with healthiness. Thus, in addition, I chose to assess whether adding a "superfood" label to food items might shift healthiness perceptions. Indeed, it was hypothesized that all three labels would increase the healthiness ratings of a given food relative to a control condition. Furthermore, if associations were made more salient with the use of these three labels, it would also be expected that the contribution of associations in the Vector Representation Model would further increase relative to the Nutrient Model.

## Methods

### *Design and Procedure*

In the same manner as the studies from Chapter 3, participants were asked to judge the healthiness of various food items one at a time using a scale ranging from -100 (extremely unhealthy) to +100 (extremely healthy). I tested the effects of three non-nutrient-related labels in one study with four conditions. In the control condition, participants saw food names alongside a food image (as in Study 1C). In the three treatment conditions, participants also saw food names alongside a food image but with a non-nutrient-related label appended to the top right of the image showing either the word "organic", "fairtrade", or "superfood".

The study used the same food stimuli as Studies 1A-2C (excluding "kit kat", "skittles", "cheerios", and "corn flakes" as they are brand names) resulting in 168 food items. Participants were randomly allocated to one of four surveys and were shown all food items once in a randomized order. However, in each survey, participants were shown a quarter of foods from each condition (48 food stimuli with an organic label, 48 with a fairtrade label, 48 with a superfood label, and 48 with no label). To give an example, participants were always shown "avocado" with a "superfood" label and "broccoli" with no label in the first survey; "avocado" with an "organic" label and "broccoli" with a "fairtrade" label in the second survey; "avocado" with no label and "broccoli" with a "superfood" label in the third survey; and "avocado" with a "fairtrade" label and "broccoli" with an "organic" label in the fourth survey.

After rating all foods, participants were asked a series of questions to assess their perceptions and understanding of the non-nutrient labels used in the study. First, they were asked to rate the extent that they agreed with the statement that fairtrade, organic, or superfood products are healthier than non-fairtrade, non-organic, or non-superfood products respectively, using a 5-point Likert scale. Second, participants were asked to match the correct definition to each label name. This also included a definition of fortified food as a red herring. Finally, participants were asked about their age, gender, and dietary restrictions using the same options as in the main studies from Chapter 3 ("Pescetarian (no meat, but eat fish and/or shellfish)", "Vegetarian", "Vegan", "Other (please specify if you wish)" and "None of the above").

### ***Mixed Effect Model***

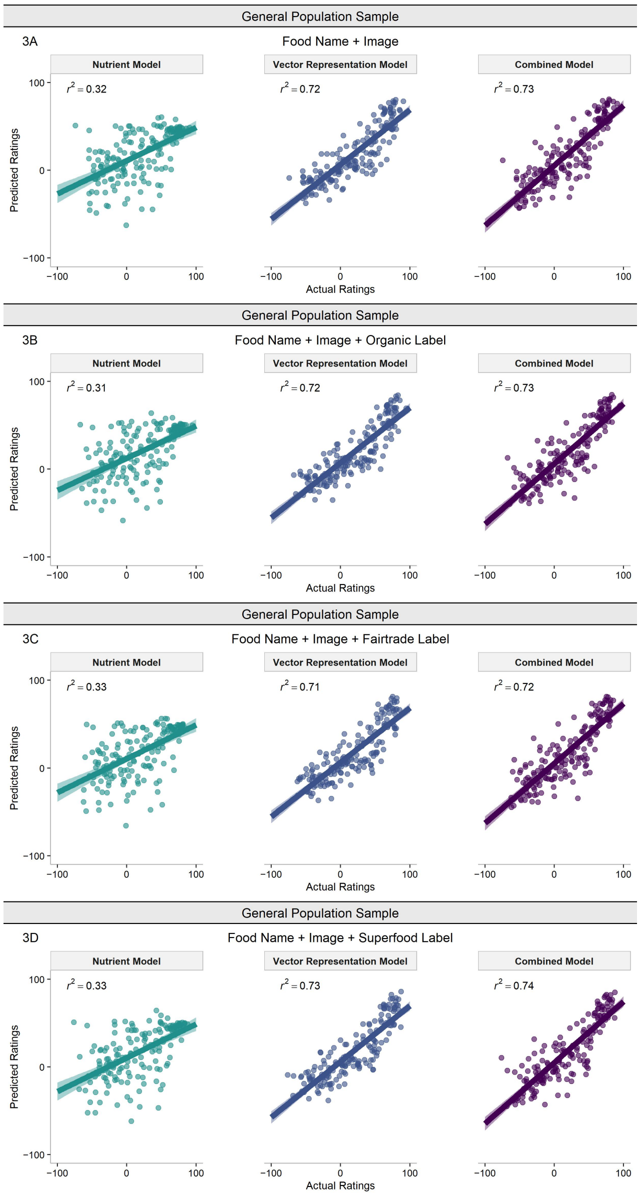
I used mixed-effects modeling to ascertain whether there was a difference between the label types of "organic", "fairtrade", "superfood", and no label (control) in the ability to shift people's healthiness judgments of the food stimuli. This method was able to model the random effects, from participant-level differences and food stimuli differences, together with the fixed effect of label type on healthiness ratings. I chose a counterbalanced design as all participants rated all food stimuli but each participant responded to each food item in just one condition (Judd et al., 2017). The data was analyzed using the R package "afex" (Singmann et al., 2015), which automatically loads the "lmer4" package (Bates, Maechler, Bolker, & Walker, 2015), and computes p-values via the R package "lmerTest" (Kuznetsova, Brockhoff, & Christensen, 2014). To note, p-values were calculated using the Satterthwaite approximation (Satterthwaite, 1946), a method where the degrees of freedom are lowered in an attempt to avoid Type 1 errors.

## Results

Figure 5.2 shows the high predictive ability of the Vector Representation Model in all conditions, consistent with all previous studies and conditions. The out-of-sample predictive accuracy of the Vector Representation Model was highest in the superfood condition ($r^2$ = 0.73, 95% CI = [0.68, 0.79]), followed closely by the organic condition ($r^2$ = 0.72, 95% CI = [0.67, 0.78]), control (no label) condition ($r^2$ = 0.72, 95% CI = [0.66, 0.78]), and fairtrade condition ($r^2$ = 0.71, 95% CI = [0.66, 0.77]). Interestingly, the order of the Nutrient Model's predictive ability did not follow the same pattern. The highest out-of-sample predictive accuracy was in the fairtrade condition and superfood condition, where the Nutrient Model performed identically to two decimal places ($r^2$ = 0.33, 95% CI = [0.22, 0.44]). Still, the performance of the Nutrient Model does drop slightly in the control condition ($r^2$ = 0.32, 95% CI = [0.22, 0.43]) and organic condition respectively ($r^2$ = 0.31, 95% CI = [0.20, 0.42]). Nonetheless, as expected, these results show that the predictive accuracy based on associations was higher than when based on factual nutritional content in all conditions. Moreover, akin to the pattern seen in the main studies from Chapter 3, the Combined Model has a slightly higher or equal predictive ability to the Vector Representation Model in all conditions ranging from 0.72 (95% CI [0.66, 0.78]) in the fairtrade condition to 0.74 (95% CI [0.68, 0.80]) in the superfood condition.

**Figure 5.2**

A Comparison of Predictive Accuracy between Models in Studies 3A, 3B, 3C, and 3D




*Note.* A comparison of predictive accuracy between models that used only nutrient content, only word vector representations, or a combination of nutrient content and word vector representations. Actual ratings were all from a general population sample.

Following this analysis, I performed mixed-effects modeling to establish whether any labeling type significantly affected healthiness judgments. Results showed that organic labeling significantly predicted healthiness ratings ($\beta$ = 2.074, *SE* = 0.709, *t* = 2.926, *p* = .004). However, it cannot be concluded that the fairtrade ($\beta$ = -0.331, *SE* = 0.617, *t* = -0.536, *p* = .59) or superfood label ($\beta$ = -0.809, *SE* = 0.654, *t* = -1.236, *p* = .22) significantly affected healthiness ratings in this study. However, when using the "r.squaredGLMM" function within the "MuMIn" R package, it was found that while 62% of the variance in healthiness ratings could be explained by the whole model, only .03% could be explained by the fixed factors of label type. Therefore, as further detailed in the discussion, these findings must be taken with caution.

### Discussion

The results of this study again demonstrate the contribution of using vector representations of food items to predict and explain healthiness judgments. There is a consistent pattern with the main studies from Chapter 3, where associations between individual food items and concepts relating to naturalness and rawness were a better predictor of healthiness judgments than key nutrient content in all conditions. Indeed, each of the best-fit models in all four conditions predicted very similar words when making healthiness predictions for "other" words in the English language, to each other and the main studies from Chapter 3. Nonetheless, the predictive power of each computational model is relatively similar between the labeling types and the control condition. Therefore, it is hard to distinguish the unique influence of each labeling type on judgments using only this approach.

Using mixed-effects modeling, only the organic labeling condition was found to significantly influence healthiness judgments. This finding is in line with previous research (Besson et al., 2019), which found that 46% of people considered the presence of an organic label as evidence that the product is healthy. An advantage of the approach used in this study was that it could consider whether this effect was true across a wide range of foods across food categories, rather than on a single food item or category as in previous research. Nonetheless, I should highlight that this study was only a first step in ascertaining the influence of all three labels on food packaging. The current design of the food stimuli presented to participants in this study is not the format used in naturalistic settings. I made this decision because there is no existing superfood label, as it is a term banned on food packaging. Therefore, I opted for a consistent design between the three labels in this study, where only the non-nutrient label word differed. As a result, in subsequent research, I recommend that researchers test design elements from existing designs of organic and fairtrade labels singularly to assess how these different features (e.g. color, shape) influence healthiness perceptions.

It is worth mentioning that for the mixed effects model, I started with a maximal model as recommended by Bates et al. (2015). However, I faced the commonly encountered issue of the model failing to converge, likely due to the model being too complex for the data. From following the advice outlined in Barr et al. (2013); Bates et al. (2015); Matuschek et al. (2017); Singmann and Kellen (2019), the next step was to simplify the model as minimally as possible to increase power. I first removed the random intercept only, keeping the slope, but this model still failed to converge. Second, I ran a zero-random-correlation model, where random correlations between the random intercepts and slopes are ignored, but this also led to a failure to converge. Therefore, I implemented a solution from the "afex" R package, which suppresses the correlation between factors by transforming them into numerical covariates. While this model did converge, I still received a warning for singularity. Singular fits are often indicative of overfitting (Barr et al., 2013; Matuschek et al., 2017), implying that more observations are necessary to distinguish group variation from residual variation.

There are two ways to obtain more observations, increasing participant sample size or increasing the number of food stimuli. My main intention for this study was to replicate the methodology of the main studies from Chapter 3 using a different food packaging design feature. Therefore, similar to these previous studies, I did not apply standard power analysis to calculate the sample size of participants. Instead, my estimates for sample size were informed by previous research that used a similar predictive (machine learning) methodology (Richie et al., 2019). However, while previous studies focused on analyses using aggregate-level data, I also included a mixed-effects model that considers all observations. Unfortunately, I did not take this into account and it is likely that the study was not sufficiently powered for finding a small effect size like the one expected here.

Regarding the number of food stimuli, it has since been recommended that the optimal number of food stimuli for computational modeling is between 200-300

(Bhatia et al., 2021a), as the algorithms require a fair amount of training data. I had 172 food items in the main studies from Chapter 3, which my collaborators and I reduced from 200 due to ambiguity and/or potential unfamiliarity in the UK. In the present study, I reduced the number of food stimuli to 168, as they were brand names (e.g. Kit Kat). Thus, my suggestion to future researchers would be to prioritize increasing the number of food stimuli to increase the power when conducting mixed effect models in similar studies, as this will also result in better model predictions.

### 5.2.3 Recommendations

Beyond the studies already mentioned, I encourage researchers to conduct another empirical study to test the validity of the associations found using my vector space approach. Since I collected my data, a new nutrient profiling system has been developed called the Food Compass System (Mozaffarian et al., 2021). This approach addresses many of the limitations of previous nutrient profiling systems, which rely on isolating single nutrients and cannot account for food processing methods. Instead, the Food Compass systematically scores all foods on 54 attributes (including vitamins, minerals, additives, and processing) per 100 kcal, providing a final Food Compass Score from 1 (least healthy) to 100 (most healthy). The supplemental materials for this paper contain Food Compass Scores, Health Star Ratings, Nutri-Score, and NOVA classifications for over 8,000 unique foods and beverages commonly consumed in the USA. Like in the USDA database, the descriptions of these foods contain the level of processing of each item (e.g., "Orange Juice, 100%, freshly squeezed"). In all my studies, I used the nutrient composition data for the rawest version of each food item for consistency and because I was interested in capturing associations of foods at their most abstract level. While it is possible to manipulate the wording of a food item description to find it in the word2vec vocabulary (e.g., "freshly_squeezed_orange_juice" and "steamed_broccoli" are present), this level of specificity is not available for many foods in word2vec. However, a solution may be to present images to participants of the same foods in different processing stages, which can then be used to compare shifts in perceptions due to the degree of processing. It would then be interesting to compare the predictive performance of the word2vec model against the scores provided by these different nutrient systems, to assess how well the information captured by these systems overlaps with the associations with food processing uncovered from knowledge representations in language. With these recommendations, I believe any extension of this work would produce a more complete picture of how food healthiness perceptions may shift in naturalistic settings, in line with my original ambitions.

# References

Adamski, M., Truby, H., M Klassen, K., Cowan, S., and Gibson, S. (2020). Using the internet: Nutrition information-seeking behaviours of lay people enrolled in a massive online nutrition course. *Nutrients*, 12(3):750.

Aka, A. and Bhatia, S. (2021). Machine learning models for predicting, understanding and influencing health perception.

Al-Ani, H. H., Devi, A., Eyles, H., Swinburn, B., and Vandevijvere, S. (2016). Nutrition and health claims on healthy and less-healthy packaged food products in new zealand. *British Journal of Nutrition*, 116(6):1087–1094.

Ambwani, S., Shippe, M., Gao, Z., and Austin, S. B. (2019). Is# cleaneating a healthy or harmful dietary strategy? perceptions of clean eating and associations with disordered eating among young adults. *Journal of eating disorders*, 7(1):1–14.

Amos, C., Hansen, J. C., and King, S. (2019). All-natural versus organic: are the labels equivalent in consumers' minds? *Journal of Consumer Marketing*, 36(4):516–526.

An, R., Shi, Y., Shen, J., Bullard, T., Liu, G., Yang, Q., Chen, N., and Cao, L. (2021). Effect of front-of-package nutrition labeling on food purchases: a systematic review. *Public Health*, 191:59–67.

Andrews, J. C., Burton, S., and Netemeyer, R. G. (2000). Are some comparative nutrition claims misleading? the role of nutrition knowledge, ad claim type and disclosure conditions. *Journal of Advertising*, 29(3):29–42.

Ares, G., Mawad, F., Giménez, A., and Maiche, A. (2014). Influence of rational and intuitive thinking styles on food choice: Preliminary evidence from an eye-tracking study with yogurt labels. *Food Quality and Preference*, 31:28–37.

Armstrong, B., Bridge, G., Oakden, L., Reynolds, C., Wang, C., Panzone, L. A., Rivera, X. S., Kause, A., Ffoulkes, C., Krawczyk, C., et al. (2020). Piloting citizen science methods to measure perceptions of carbon footprint and energy content of food. *Frontiers in Sustainable Food Systems*, 4:120.

Arseniev-Koehler, A. and Foster, J. G. (2020). Machine learning as a model for cultural learning: Teaching an algorithm what it means to be fat. *arXiv preprint arXiv:2003.12133*.

Asch, S. E. (1961). *Forming impressions of personality*. University of California Press.

Bandy, L. K., Hollowell, S., Harrington, R., Scarborough, P., Jebb, S., and Rayner, M. (2021). Assessing the healthiness of uk food companies' product portfolios using food sales and nutrient composition data. *PloS one*, 16(8):e0254833.

Barklamb, A. M., Molenaar, A., Brennan, L., Evans, S., Choong, J., Herron, E., Reid, M., and McCaffrey, T. A. (2020). Learning the language of social media: A comparison of engagement metrics and social media strategies used by food and nutrition-related social media accounts. *Nutrients*, 12(9):2839.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Barron, A., Huang, J., Spang, R., and DeDeo, S. (2018). Individuals, institutions, and innovation in the debates of the french revolution. *Proceedings of the National Academy of Sciences*, 115(18):4607–4612.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bauer, U. E., Briss, P. A., Goodman, R. A., and Bowman, B. A. (2014). Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the usa. *The Lancet*, 384(9937):45–52.

Bedoya-Perales, N. S., Pumi, G., Talamini, E., and Padula, A. D. (2018). The quinoa boom in Peru: Will land competition threaten sustainability in one of the cradles of agriculture? *Land Use Policy*, 79:475–480.

Benoit, K. and Matsuo, A. (2020). *spacyr: Wrapper to the 'spaCy' 'NLP' Library*. R package version 1.2.1.

Berry, C. and Romero, M. (2021). The fair trade food labeling health halo: Effects of fair trade labeling on consumption and perceived healthfulness. *Food Quality and Preference*, 94:104321.

Bharadwaj, P. and Shao, Z. (2019). Fake News Detection with Semantic Features and Text Mining. *International Journal on Natural Language Computing (IJNLC)*, 8(3).

Bhatia, N. and Bhatia, S. (2021). Changes in gender stereotypes over time: a computational analysis. *Psychology of Women Quarterly*, 45(1):106–125.

Bhatia, S. (2016). Vector space semantic models predict subjective probability judgments for real-world events. In *CogSci*.

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological review*, 124(1):1.

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

Bhatia, S., Olivola, C. Y., Bhatia, N., and Ameen, A. (2021a). Predicting leadership perception with large-scale natural language data. *The Leadership Quarterly*, page 101535.

Bhatia, S. and Stewart, N. (2018). Naturalistic multiattribute choice. *Cognition*, 179:71–88.

Bhatia, S., Walasek, L., Slovic, P., and Kunreuther, H. (2021b). The more who die, the less we care: Evidence from natural language analysis of online news articles and social media posts. *Risk Analysis*, 41(1):179–203.

Blackburn, K. G., Yilmaz, G., and Boyd, R. L. (2018). Food for thought: Exploring how people think and talk about food online. *Appetite*, 123:390–401.

Blechert, J., Lender, A., Polk, S., Busch, N. A., and Ohla, K. (2019). Food-pics_extended—an image database for experimental research on eating and appetite: additional images, normative ratings and an updated review. *Frontiers in psychology*, 10:307.

Borgmeier, I. and Westenhoefer, J. (2009). Impact of different food label formats on healthiness evaluation and food choice of consumers: a randomized-controlled study. *BMC public health*, 9(1):1–12.

Breen, M., James, H., Rangan, A., and Gemming, L. (2020). Prevalence of product claims and marketing buzzwords found on health food snack products does not relate to nutrient profile. *Nutrients*, 12(5):1513.

Brown, K. A., Timotijevic, L., Barnett, J., Shepherd, R., Lähteenmäki, L., and Raats, M. M. (2011). A review of consumer awareness, understanding and use of food-based dietary guidelines. *British Journal of Nutrition*, 106(1):15–26.

Brownbill, A. L., Braunack-Mayer, A. J., and Miller, C. L. (2020). What makes a beverage healthy? A qualitative study of young adults' conceptualisation of sugar-containing beverage healthfulness. *Appetite*, 150:104675.

Brownbill, A. L., Miller, C. L., and Braunack-Mayer, A. J. (2018). Industry use of 'better-for-you'features on labels of sugar-containing beverages. *Public health nutrition*, 21(18):3335–3343.

Bucher, T., Collins, C., Diem, S., and Siegrist, M. (2016). Adolescents' perception of the healthiness of snacks. *Food Quality and Preference*, 50:94–101.

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

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

Bui, M. M., Tangari, A. H., and Haws, K. L. (2017). Can health "halos" extend to food packaging? an investigation into food healthfulness perceptions and serving sizes on consumption decisions. *Journal of Business Research*, 75:221–228.

Bullinaria, J. A. and Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods*, 39(3):510–526.

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2020). The structure of economic news. Working Paper 26648, National Bureau of Economic Research.

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

Campos, S., Doxey, J., and Hammond, D. (2011). Nutrition labels on pre-packaged foods: a systematic review. *Public health nutrition*, 14(8):1496–1506.

Cancer Research UK (2020). Food myths.

Carels, R. A., Harper, J., and Konrad, K. (2006). Qualitative perceptions and caloric estimations of healthy and unhealthy foods by behavioral weight loss participants. *Appetite*, 46(2):199–206.

Carels, R. A., Konrad, K., and Harper, J. (2007). Individual differences in food perceptions and calorie estimation: an examination of dieting status, weight, and gender. *Appetite*, 49(2):450–458.

Carlsson, F., Kataria, M., Lampi, E., Nyberg, E., and Sterner, T. (2021). Red, yellow, or green? do consumers' choices of food products depend on the label design? *European Review of Agricultural Economics*.

Cecchini, M. and 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.

Chandon, P. and Wansink, B. (2007). The biasing health halos of fast-food restaurant health claims: lower calorie estimates and higher side-dish consumption intentions. *Journal of Consumer Research*, 34(3):301–314.

Chandon, P. and Wansink, B. (2012). Does food marketing need to make us fat? a review and solutions. *Nutrition reviews*, 70(10):571–593.

Chapman, G. and Maclean, H. (1993). "junk food" and "healthy food": meanings of food in adolescent women's culture. *Journal of nutrition education*, 25(3):108–113.

Charlesworth, T. E. and Banaji, M. R. (2021). Word embeddings reveal social group attitudes and stereotypes in large language corpora. *Atlas of Language Analysis in Psychology*.

Cloutier, K., Mongeau, L., Pageau, M., and Provencher, V. (2013). Food perceptions among adults and registered dietitians: Are they similar? *Food and Nutrition*, 4(10A):2–8.

Coates, A. E., Hardman, C. A., Halford, J. C. G., Christiansen, P., and Boyland, E. J. (2020). "it's just addictive people that make addictive videos": Children's understanding of and attitudes towards influencer marketing of food and beverages by youtube video bloggers. *International journal of environmental research and public health*, 17(2):449.

Commission, E. (2011). Regulation (eu) no 1169/2011 of the european parliament and of the council of 25 october 2011 on the provision of food information to consumers, amending regulations (ec) no 1924/2006 and (ec) no 1925/2006 of the european parliament and of the council, and repealing commission directive 87/250/eec, council directive 90/496/eec, commission directive 1999/10/ec, directive 2000/13/ec of the european parliament and of the council, commission directives 2002/67/ec and 2008/5/ec and commission regulation (ec) no 608/2004. *Official Journal of the European Commission*, 304:18—63.

Cowburn, G. and Stockley, L. (2005). Consumer understanding and use of nutrition labelling: a systematic review. *Public health nutrition*, 8(1):21–28.

Croll, J. K., Neumark-Sztainer, D., and Story, M. (2001). Healthy eating: what does it mean to adolescents? *Journal of nutrition education*, 33(4):193–198.

Curll, J., Parker, C., MacGregor, C., and Petersen, A. (2016). Unlocking the energy of the amazon? the need for a food fraud policy approach to the regulation of anti-ageing health claims on superfood labelling. *Federal Law Review*, 44(3):419–449.

Davidović, D., Paunović, K., Zarić, D., Jovanović, A., Vasiljević, N., Stošović, D., and Tomanić, M. (2021). Nutrition and health claims spectra of pre-packaged foods on serbian supermarket shelves: a repeated cross-sectional study. *Nutrients*, 13(8):2832.

Davies, M. (2017). The new 4.3 billion word now corpus, with 4–5 million words of data added every day. In *The 9th International Corpus Linguistics Conference*.

Davies, M., Hegedűs, I., and Fodor, A. (2012). The 400 million word corpus of historical american english (1810–2009). In *English Historical Linguistics 2010: Selected Papers from the Sixteenth International Conference on English Historical Linguistics (ICEHL 16), Pécs, 23-27 August 2010*, volume 325, page 231. John Benjamins Publishing.

De Deyne, S., Perfors, A., and 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*, pages 1861–1870.

de Ridder, D., Kroese, F., Evers, C., Adriaanse, M., and Gillebaart, M. (2017). Healthy diet: Health impact, prevalence, correlates, and interventions. *Psychology & health*, 32(8):907–941.

de Rooij, M. and Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, 3(2):248–263.

De Vlieger, N. M., Collins, C., and Bucher, T. (2017). What is a nutritious snack? level of processing and macronutrient content influences young adults' perceptions. *Appetite*, 114:55–63.

Delgado, A., Issaoui, M., Vieira, M. C., Saraiva de Carvalho, I., and Fardet, A. (2021). Food composition databases: Does it matter to human health? *Nutrients*, 13(8):2816.

Delicato, C., Salvatore, F. P., and Contò, F. (2019). Consumers' understanding of healthy foods: The evidence on superfoods in belgium. 2019 Eighth AIEAA Conference, June 13-14, Pistoia, Italy 300907, Italian Association of Agricultural and Applied Economics (AIEAA).

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Douma, E. R., Valente, M., and Syurina, E. V. (2021). Developmental pathway of orthorexia nervosa: Factors contributing to progression from healthy eating to excessive preoccupation with healthy eating. experiences of dutch health professionals. *Appetite*, 158:105008.

Downs, J. S., Loewenstein, G., and Wisdom, J. (2009). Strategies for promoting healthier food choices. *American Economic Review*, 99(2):159–64.

Downs, J. S., Wisdom, J., and Loewenstein, G. (2015). Helping consumers use nutrition information: Effects of format and presentation. *American Journal of Health Economics*, 1(3):326–344.

Ettlin, F., Bröder, A., and Henninger, M. (2015). A new task format for investigating information search and organization in multiattribute decisions. *Behavior research methods*, 47(2):506–518.

European Commission (2006). Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods. *Official Journal of the European Union*, L 404:9–25.

European Food Information Council (2016). Global update on nutrition labelling.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59:255–278.

Fernan, C., Schuldt, J. P., and Niederdeppe, J. (2018). Health halo effects from product titles and nutrient content claims in the context of "protein" bars. *Health communication*, 33(12):1425–1433.

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., and 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.

Firth, J. R. (1957). Papers in linguistic 1934–1951, london: Oxford univ.

Food Standards Australia New Zealand (2018). Australia new zealand food standards code – standard 1.2.7 – nutrition, health and related claims.

Foroni, F., Esmaeilikia, M., and Rumiati, R. I. (2022). What makes a food healthy? sex differences in what is associated to healthiness evaluations. *Food Quality and Preference*, 96:104438.

Forouhi, N. G., Krauss, R. M., Taubes, G., and Willett, W. (2018). Dietary fat and cardiometabolic health: evidence, controversies, and consensus for guidance. *Bmj*, 361.

Franco Lucas, B., Costa, J. A. V., and Brunner, T. A. (2021). Superfoods: Drivers for consumption. *Journal of Food Products Marketing*, 27(1):1–9.

Frederick, D. A., Tomiyama, A. J., Bold, J. G., and Saguy, A. C. (2020). Can she be healthy at her weight? effects of news media frames on antifat attitudes, dieting intentions, and perceived health risks of obesity. *Stigma and Health*, 5(3):247–257.

Freeland-Graves, J. H. and Nitzke, S. (2013). Position of the academy of nutrition and dietetics: total diet approach to healthy eating. *Journal of the Academy of Nutrition and Dietetics*, 113(2):307–317.

Galfano, V., Syurina, E. V., Valente, M., and Donini, L. M. (2022). When "healthy" is taken too far: Orthorexia nervosa—current state, controversies and future directions. In Manzato, E., Cuzzolaro, M., and Donini, L. M., editors, *Hidden* *and Lesser-known Disordered Eating Behaviors in Medical and Psychiatric Conditions*, pages 159–176. Springer International Publishing.

Garcia, C. J. and Proffitt, J. M. (2021). "We are Coca-Cola and so much more": political economic analysis of non-carbonated SSB Coke brands. *Food, Culture & Society*, pages 1–18.

Gaspar, M. C. d. M. P., Garcia, A. M., and Larrea-Killinger, C. (2020). How would you define healthy food? social representations of brazilian, french and spanish dietitians and young laywomen. *Appetite*, 153:104728.

Gefen, D., Endicott, J. E., Fresneda, J. E., Miller, J., and Larsen, K. R. (2017). A guide to text analysis with latent semantic analysis in r with annotated code: Studying online reviews and the stack exchange community. *Communications of the Association for Information Systems*, 41(1):21.

Gibson-Moore, H. and Spiro, A. (2021). Evolution not revolution – what might the future hold for front-of-pack nutrition labelling in the uk?: A british nutrition foundation roundtable. *Nutrition Bulletin*, 46(3):383–394.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62:451–482.

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

Goldberg, J. P. and Sliwa, S. A. (2011). Communicating actionable nutrition messages: challenges and opportunities. *Proceedings of the Nutrition Society*, 70(1):26–37.

Gomez, P. (2013). Common biases and heuristics in nutritional quality judgments: a qualitative exploration. *International Journal of Consumer Studies*, 37(2):152–158.

González-Vallejo, C., Lavins, B. D., and Carter, K. A. (2016). Analysis of nutrition judgments using the nutrition facts panel. *Appetite*, 105:71–84.

Graeff-Hönninger, S. and Khajehei, F. (2019). The demand for superfoods: Consumers' desire, production viability and bio-intelligent transition. In Piatti, C., Graeff-Hönninger, S., and Khajehei, F., editors, *Food Tech Transitions: Reconnecting Agri-Food, Technology and Society*, pages 81–94. Springer International Publishing.

Grand, G., Blank, I., Pereira, F., and Fedorenko, E. (2018). Semantic projection: Predicting high-level judgment recovering human knowledge of multiple, distinct object features from word embeddings. *arXiv preprint arXiv*.

Green, H. (2015). Should foods or nutrients be the focus of guidelines to promote healthful eating? *Nutrition Bulletin*, 40(4):296–302.

Gressier, M., Sassi, F., and Frost, G. (2020). Healthy foods and healthy diets. how government policies can steer food reformulation. *Nutrients*, 12(7):1992.

Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological review*, 114(2):211.

Groeniger, J. O., van Lenthe, F. J., Beenackers, M. A., and Kamphuis, C. B. (2017). Does social distinction contribute to socioeconomic inequalities in diet: the case of 'superfoods' consumption. *International Journal of Behavioral Nutrition and Physical Activity*, 14(1):1–7.

Grunert, K. G. and Wills, J. M. (2007). A review of european research on consumer response to nutrition information on food labels. *Journal of public health*, 15(5):385–399.

Günther, F., Rinaldi, L., and Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, 14(6):1006–1033.

Haasova, S. and Florack, A. (2019). Practicing the (un) healthy= tasty intuition: Toward an ecological view of the relationship between health and taste in consumer judgments. *Food Quality and Preference*, 75:39–53.

Hackman, E. M. and Moe, G. L. (1999). Evaluation of newspaper reports of nutrition-related research. *Journal of the American Dietetic Association*, 99(12):1564–1566.

Hagmann, D. and Siegrist, M. (2020). Nutri-score, multiple traffic light and incomplete nutrition labelling on food packages: Effects on consumers' accuracy in identifying healthier snack options. *Food Quality and Preference*, 83:103894.

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3):146–162.

Hills, T. T., Proto, E., Sgroi, D., and Seresinhe, C. I. (2019). Historical analysis of national subjective wellbeing using millions of digitized books. *Nature human behaviour*, 3(12):1271–1275.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8):1–28.

Huang, L. and Lu, J. (2016). The impact of package color and the nutrition content labels on the perception of food healthiness and purchase intention. *Journal of food products marketing*, 22(2):191–218.

Inoue-Choi, M., Oppeneer, S. J., and Robien, K. (2013). Reality check: there is no such thing as a miracle food. *Nutrition and cancer*, 65(2):165–168.

Jagdale, Y. D., Mahale, S. V., Zohra, B., Nayik, G. A., Dar, A. H., Khan, K. A., Abdi, G., and Karabagias, I. K. (2021). Nutritional profile and potential health benefits of super foods: A review. *Sustainability*, 13(16):9240.

Jones, A., Neal, B., Reeve, B., Mhurchu, C. N., and Thow, A. M. (2019). Front-of-pack nutrition labelling to promote healthier diets: current practice and opportunities to strengthen regulation worldwide. *BMJ global health*, 4(6):e001882.

Jones, M. N., Gruenenfelder, T. M., and Recchia, G. (2018). In defense of spatial models of semantic representation. *New Ideas in Psychology*, 50:54–60.

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

Judd, C. M., Westfall, J., and Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual review of psychology*, 68(1):601–625.

Julia, C., Fialon, M., Galan, P., Deschasaux-Tanguy, M., Andreeva, V. A., Kesse-Guyot, E., Touvier, M., and Hercberg, S. (2021). Are foods 'healthy'or 'healthier'? front-of-pack labelling and the concept of healthiness applied to foods. *British Journal of Nutrition*, pages 1–5.

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobel prize lecture*, 8:351–401.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 93(5):1449–1475.

Kahneman, D. and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49:81.

Kanter, R., Vanderlee, L., and Vandevijvere, S. (2018). Front-of-package nutrition labelling policy: global progress and future directions. *Public Health Nutrition*, 21(8):1399–1408.

Kaur, A., Scarborough, P., Matthews, A., Payne, S., Mizdrak, A., and Rayner, M. (2016). How many foods in the uk carry health and nutrition claims, and are they healthier than those that do not? *Public health nutrition*, 19(6):988–997.

Kissock, K. R., Vieux, F., Mathias, K. C., Drewnowski, A., Seal, C. J., Masset, G., Smith, J., Mejborn, H., McKeown, N. M., and Beck, E. J. (2021). Aligning nutrient profiling with dietary guidelines: modifying the nutri-score algorithm to include whole grains. *European journal of nutrition*, pages 1–13.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., and Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4):150.

Kāle, M. and Agbozo, E. (2020). Healthy food depiction on social media: The case of kale on twitter. *CEUR Workshop Proceedings*, 2865:51–62.

Ladher, N. (2016). Nutrition science in the media: you are what you read. *BMJ*, 353.

Laguna-Camacho, A. and Booth, D. A. (2015). Meals described as healthy or unhealthy match public health education in england. *Appetite*, 87:283–287.

Laguna-Camacho, A., García-Manjarrez, E., Frayn, M., Knäuper, B., Domínguez-García, M. V., and Escalante-Izeta, E. I. (2018). Perceived healthiness of breakfasts in women with overweight or obesity match expert recommendations. *Open Psychology*, 1(1):25–35.

Landauer, T. K. and 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.

Larkin, D. and Martin, C. R. (2016). Caloric estimation of healthy and unhealthy foods in normal-weight, overweight and obese participants. *Eating behaviors*, 23:91–96.

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

Li, Y., Hills, T., and Hertwig, R. (2020). A brief history of risk. *Cognition*, 203:104344.

Lin, Y., Michel, J.-B., Lieberman, E. A., Orwant, J., Brockman, W., and Petrov, S. (2012). Syntactic annotations for the google books ngram corpus. In *Proceedings of the ACL 2012 system demonstrations*, pages 169–174.

Liu, H., Meng-Lewis, Y., Ibrahim, F., and Zhu, X. (2021). Superfoods, super healthy: Myth or reality? examining consumers' repurchase and wom intention regarding superfoods: A theory of consumption values perspective. *Journal of Business Research*, 137:69–88.

Lobstein, T. and Davies, S. (2009). Defining and labelling 'healthy' and 'unhealthy' food. *Public health nutrition*, 12(3):331–340.

Lobstein, T., Landon, J., and Lincoln, P. (2007). Misconceptions and misinformation: the problems with guideline daily amounts (gdas). In *A Review of GDAs and Their Use for Signalling Nutritional Information on Food and Drink Labels. London: National Heart Forum.*

Loyer, J. (2016). *The social lives of superfoods.* PhD thesis, University of Adelaide, South Australia.

Lusk, J. L. (2019). Consumer beliefs about healthy foods and diets. *PLoS One*, 14(10):e0223098.

Lynn, T., Rosati, P., Leoni Santos, G., and Endo, P. T. (2020). Sorting the healthy diet signal from the social media expert noise: Preliminary evidence from the healthy diet discourse on twitter. *International Journal of Environmental Research and Public Health*, 17(22):8557.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., and Carney, J. (2019). The lancaster sensorimotor norms: multidimensional measures of perceptual and action strength for 40,000 english words. *Behavior Research Methods*, pages 1–21.

MacGregor, C., Petersen, A., and Parker, C. (2021). Promoting a healthier, younger you: the media marketing of anti-ageing superfoods. *Journal of Consumer Culture*, 21(2):164–179.

Magrach, A. and Sanz, M. J. (2020). Environmental and social consequences of the increase in the demand for ‘superfoods’ world-wide. *People and Nature*, 2(2):267–278.

Manippa, V., Giuliani, F., and Brancucci, A. (2020). Healthiness or calories? side biases in food perception and preference. *Appetite*, 147:104552.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language*, 94:305–315.

Mazzocchi, M., Cagnone, S., Bech-Larsen, T., Niedźwiedzka, B., Saba, A., Shankar, B., Verbeke, W., and Traill, W. B. (2015). What is the public appetite for healthy eating policies? evidence from a cross-european survey. *Health Economics, Policy and Law*, 10(3):267–292.

McCandless, D., Quick, M., Starling, S., and Geere, D. (2018). Information is beautiful: Snake oil superfoods.

Meyerding, S. G., Kürzdörfer, A., and Gassler, B. (2018). Consumer preferences for superfood ingredients—the case of bread in germany. *Sustainability*, 10(12):4667.

Mhurchu, C. N., Eyles, H., Jiang, Y., and Blakely, T. (2018). Do nutrition labels influence healthier food choices? analysis of label viewing behaviour and subsequent food purchases in a labelling intervention trial. *Appetite*, 121:360–365.

Michel, F., Sanchez-Siles, L. M., and Siegrist, M. (2021). Predicting how consumers perceive the naturalness of snacks: The usefulness of a simple index. *Food Quality and Preference*, 94:104295.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., and Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.

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

Miller, L. M. S. and Cassady, D. L. (2015). The effects of nutrition knowledge on food label use. a review of the literature. *Appetite*, 92:207–216.

Mintel (2016). Super growth for “super” foods: New product development shoots up 202% globally over the past five years.

Morewedge, C. K. and Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in cognitive sciences*, 14(10):435–440.

Motoki, K., Park, J., Pathak, A., and Spence, C. (2021). Constructing healthy food names: On the sound symbolism of healthy food. *Food Quality and Preference*, 90:104157.

Mötteli, S., Keller, C., Siegrist, M., Barbey, J., and Bucher, T. (2016). Consumers' practical understanding of healthy food choices: a fake food experiment. *British Journal of Nutrition*, 116(3):559–566.

Mozaffarian, D., El-Abbadi, N. H., O'Hearn, M., Erndt-Marino, J., Masters, W. A., Jacques, P., Shi, P., Blumberg, J. B., and Micha, R. (2021). Food compass is a nutrient profiling system using expanded characteristics for assessing healthfulness of foods. *Nature Food*, 2(10):809–818.

Muller, L. and Ruffieux, B. (2020). What makes a front-of-pack nutritional labelling system effective: the impact of key design components on food purchases. *Nutrients*, 12(9):2870.

Nagler, R. H. (2014). Adverse outcomes associated with media exposure to contradictory nutrition messages. *Journal of health communication*, 19(1):24–40.

Nicholes, M. J., Quested, T. E., Reynolds, C., Gillick, S., and Parry, A. D. (2019). Surely you don't eat parsnip skins? categorising the edibility of food waste. *Resources, Conservation and Recycling*, 147:179–188.

Nicklas, T. A., Drewnowski, A., and O'Neil, C. E. (2014). The nutrient density approach to healthy eating: challenges and opportunities. *Public health nutrition*, 17(12):2626–2636.

Nicklas, T. A., O'Neil, C. E., Mendoza, J., Liu, Y., Zakeri, I. F., and Berenson, G. S. (2008). Are energy dense diets also nutrient dense? *Journal of the American College of Nutrition*, 27(5):553–560.

Oakes, M. E. (2004). *Bad foods: Changing attitudes about what we eat.* Routledge, 1st edition.

Oakes, M. E. and Slotterback, C. S. (2001). Gender differences in perceptions of the healthiness of foods. *Psychology and Health*, 16(1):57–65.

Oakes, M. E. and Slotterback, C. S. (2002). The good, the bad, and the ugly: Characteristics used by young, middle-aged, and older men and women, dieters and non-dieters to judge healthfulness of foods. *Appetite*, 38(2):91–97.

Ogundijo, D., Tas, A., and Onarinde, B. (2021). An assessment of nutrition information on front of pack labels and healthiness of foods in the united kingdom retail market. *BMC Public Health*, 21(1):1–10.

Oljans, E., Elmståhl, H., Mattsson Sydner, Y., and Hjälmeskog, K. (2018). From nutrients to wellbeing identifying discourses of food in relation to health in syllabi. *Pedagogy, Culture & Society*, 26(1):35–49.

Ooms, J. (2020). *hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker*. R package version 3.0.1.

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

Orquin, J. L. and Scholderer, J. (2015). Consumer judgments of explicit and implied health claims on foods: Misguided but not misled. *Food policy*, 51:144–157.

Pandur, M. B., Dobša, J., and Kronegger, L. (2020). Topic modelling in social sciences-case study of web of science. In *Central European Conference on Information and Intelligent Systems*, pages 211–218. Faculty of Organization and Informatics Varazdin.

Papies, E. K. (2013). Tempting food words activate eating simulations. *Frontiers in psychology*, 4:838.

Papies, E. K., Barsalou, L. W., and Rusz, D. (2020). Understanding desire for food and drink: a grounded-cognition approach. *Current Directions in Psychological Science*, 29(2):193–198.

Paquette, M.-C. (2005). Perceptions of healthy eating: state of knowledge and research gaps. *Canadian Journal of Public Health*, 96(3):S16–S21.

Parker, C., Johnson, H., and Curll, J. (2019). Consumer power to change the food system: a critical reading of food labels as governance spaces: the case of acai berry superfoods. *J. Food L. & Pol'y*, 15:1.

Patel, A., Sands, A., Callison-Burch, C., and Apidianaki, M. (2018). Magnitude: A fast, efficient universal vector embedding utility package. *arXiv preprint arXiv:1810.11190*.

Payne, J. W., Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge university press.

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

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

Perkovic, S. and Orquin, J. L. (2018). Implicit statistical learning in real-world environments leads to ecologically rational decision making. *Psychological Science*, 29(1):34–44.

Perkovic, S., Otterbring, T., Schärli, C., and Pachur, T. (2021). The perception of food products in adolescents, lay adults, and experts: A psychometric approach. *Journal of Experimental Psychology: Applied*.

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. arxiv 2018. *arXiv preprint arXiv:1802.05365*, 12.

Pieniak, Z., Verbeke, W., Vanhonacker, F., Guerrero, L., and Hersleth, M. (2009). Association between traditional food consumption and motives for food choice in six european countries. *Appetite*, 53(1):101–108.

Pilař, L., Kvasničková Stanislavská, L., and Kvasnička, R. (2021). Healthy food on the twitter social network: Vegan, homemade, and organic food. *International Journal of Environmental Research and Public Health*, 18(7):3815.

Pilgrim, K. and Bohnet-Joschko, S. (2019). Selling health and happiness how influencers communicate on instagram about dieting and exercise: Mixed methods research. *BMC Public Health*, 19(1):1–9.

Pinho-Gomes, A.-C., Kaur, A., Scarborough, P., and Rayner, M. (2021). Are the eatwell guide and nutrient profiling models consistent in the uk? *Nutrients*, 13(8):2732.

Plasek, B., Lakner, Z., and Temesi, Á. (2020). Factors that Influence the Perceived Healthiness of Food—Review. *Nutrients*, 12(6):1881.

Plasek, B., Lakner, Z., and Temesi, Á. (2021). I believe it is healthy—impact of extrinsic product attributes in demonstrating healthiness of functional food products. *Nutrients*, 13(10):3518.

Proestos, C. (2018). Superfoods: Recent data on their role in the prevention of diseases. *Current Research in Nutrition and Food Science Journal*, 6(3):576–593.

Provencher, V., Polivy, J., and Herman, C. P. (2009). Perceived healthiness of food. if it's healthy, you can eat more! *Appetite*, 52(2):340–344.

Public Health England (2021). Mccance and widdowson's composition of foods integrated dataset.

Qutteina, Y., Hallez, L., Mennes, N., De Backer, C., and Smits, T. (2019). What do adolescents see on social media? a diary study of food marketing images on social media. *Frontiers in psychology*, 10:2637.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Raghunathan, R., Naylor, R. W., and Hoyer, W. D. (2006). The unhealthy= tasty intuition and its effects on taste inferences, enjoyment, and choice of food products. *Journal of Marketing*, 70(4):170–184.

Ramachandran, D., Kite, J., Vassallo, A. J., Chau, J. Y., Partridge, S., Freeman, B., and Gill, T. (2018). Food trends and popular nutrition advice online–implications for public health. *Online journal of public health informatics*, 10(2).

Rayner, M. (2017). Nutrient profiling for regulatory purposes. *Proceedings of the Nutrition Society*, 76(3):230–236.

Rayner, M., Scarborough, P., and Williams, C. (2004). The origin of guideline daily amounts and the food standards agency's guidance on what counts as 'a lot'and 'a little'. *Public Health Nutrition*, 7(4):549–556.

Řehůřek, R., Sojka, P., et al. (2011). Gensim—statistical semantics in python. *Retrieved from genism. org.*

Reisman, E. (2020). Superfood as spatial fix: the ascent of the almond. *Agriculture and Human Values*, 37(2):337–351.

Reutner, L., Genschow, O., and Wänke, M. (2015). The adaptive eater: Perceived healthiness moderates the effect of the color red on consumption. *Food quality and Preference*, 44:172–178.

Richie, R. and Bhatia, S. (2020). Knowledge, cognition, and everyday judgment: An introduction to the distributed semantics approach.

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

Rizk, M. T. and 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.

Roberts, M., Stewart, B., and Airoldi, E. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515):988–1003.

Roberts, M., Stewart, B., and Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, 91(1):1–40.

Roberts, M., Stewart, B., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S., and Albertson, B. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4):1064–1082.

Rodney, A. (2018). Pathogenic or health-promoting? how food is framed in healthy living media for women. *Social Science & Medicine*, 213:37–44.

Rojas-Rivas, E., Espinoza-Ortega, A., Thomé-Ortiz, H., and Moctezuma-Pérez, S. (2019). Consumers' perception of amaranth in mexico: A traditional food with characteristics of functional foods. *British Food Journal*.

Roman, S., Sánchez-Siles, L. M., and Siegrist, M. (2017). The importance of food naturalness for consumers: Results of a systematic review. *Trends in food science & technology*, 67:44–57.

Rønnow, H. N. (2020). The effect of front-of-pack nutritional labels and back-of-pack tables on dietary quality. *Nutrients*, 12(6):1704.

Ronteltap, A., Sijtsema, S. J., Dagevos, H., and de Winter, M. A. (2012). Construal levels of healthy eating. exploring consumers' interpretation of health in the food context. *Appetite*, 59(2):333–340.

Roth, A. and Zawadzki, T. (2018). Instagram as a tool for promoting superfood products. *Annals of Marketing Management & Economics*, 4(1):101–113.

Rousseau, S. (2015). The celebrity quick-fix: When good food meets bad science. *Food, Culture & Society*, 18(2):265–287.

Rubin, D. and Rosenbaum, P. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Saguy, A. C. and Almeling, R. (2008). Fat in the fire? science, the news media, and the "obesity epidemic"2. *Sociological Forum*, 23:53–83.

Šamec, D., Urlić, B., and Salopek-Sondi, B. (2019). Kale (brassica oleracea var. acephala) as a superfood: Review of the scientific evidence behind the statement. *Critical reviews in food science and nutrition*, 59(15):2411–2422.

Samoggia, A., Riedel, B., and Ruggeri, A. (2020). Social media exploration for understanding food product attributes perception: the case of coffee and health with twitter data. *British Food Journal*, 122(12):3815–3835.

Sanchez-Siles, L. M., Michel, F., Román, S., Bernal, M. J., Philipsen, B., Haro, J. F., Bodenstab, S., and Siegrist, M. (2019). The food naturalness index (fni): An integrative tool to measure the degree of food naturalness. *Trends in Food Science & Technology*, 91:681–690.

Sanjari, S. S., Jahn, S., and Boztug, Y. (2017). Dual-process theory and consumer response to front-of-package nutrition label formats. *Nutrition reviews*, 75(11):871–882.

Santos, M., Rito, A. I., Matias, F. N., Assunção, R., Castanheira, I., and Loureiro, I. (2021). Nutrient profile models a useful tool to facilitate healthier food choices: A comprehensive review. *Trends in Food Science & Technology*, 110:120–131.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6):110–114.

Scarborough, P., Boxer, A., Rayner, M., and Stockley, L. (2007a). Testing nutrient profile models using data from a survey of nutrition professionals. *Public health nutrition*, 10(4):337–345.

Scarborough, P. and Rayner, M. (2014). When nutrient profiling can (and cannot) be useful. *Public health nutrition*, 17(12):2637–2640.

Scarborough, P., Rayner, M., Stockley, L., and Black, A. (2007b). Nutrition professionals' perception of the 'healthiness' of individual foods. *Public health nutrition*, 10(4):346–353.

Scheibehenne, B., Miesler, L., and Todd, P. M. (2007). Fast and frugal food choices: Uncovering individual decision heuristics. *Appetite*, 49(3):578–589.

Schuldt, J. P. (2013). Does green mean healthy? nutrition label color affects perceptions of healthfulness. *Health communication*, 28(8):814–821.

Schuldt, J. P., Muller, D., and Schwarz, N. (2012). The "fair trade" effect: Health halos from social ethics claims. *Social Psychological and Personality Science*, 3(5):581–589.

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

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

Scrinis, G. (2013). *Nutritionism*. Columbia University Press.

Shah, A. K. and Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychological bulletin*, 134(2):207.

Shah, D., Isah, H., and Zulkernine, F. (2018). Predicting the effects of news sentiments on the stock market. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 4705–4708. IEEE.

Siipi, H. (2013). Is natural food healthy? *Journal of agricultural and environmental ethics*, 26(4):797–812.

Sikka, T. (2019). The contradictions of a superfood consumerism in a postfeminist, neoliberal world. *Food, Culture & Society*, 22(3):354–375.

Silchenko, K., Askegaard, S., and Cedrola, E. (2020). Three decades of Research in Health and Food Marketing: A Systematic Review. *Journal of Consumer Affairs*, 54(2):541–580.

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5):1–13.

Singh, M., Richie, R., and Bhatia, S. (2020). Representing and predicting everyday behavior.

Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2015). Package 'afex'.

Singmann, H. and Kellen, D. (2019). An introduction to mixed models for experimental psychology. In *New methods in cognitive psychology*, pages 4–31. Routledge.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1):3.

Song, J., Brown, M. K., Tan, M., MacGregor, G. A., Webster, J., Campbell, N. R., Trieu, K., Ni Mhurchu, C., Cobb, L. K., and He, F. J. (2021). Impact of color-coded and warning nutrition labelling schemes: A systematic review and network meta-analysis. *PLoS medicine*, 18(10):e1003765.

Spiro, A. and Wood, V. (2021). Can the concept of nutrient density be useful in helping consumers make informed and healthier food choices? a mixed-method exploratory approach.

Steinhauser, J. and 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.

Štepec, D., Tavčar, G., and Ponikvar-Svet, M. (2020). Surprisingly high fluorine content in some exotic superfoods. *Journal of Fluorine Chemistry*, 234:109521.

Steptoe, A., Pollard, T. M., and Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, 25(3):267–284.

Sütterlin, B. and Siegrist, M. (2015). Simply adding the word "fruit" makes sugar healthier: The misleading effect of symbolic information on the perceived healthiness of food. *Appetite*, 95:252–261.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of applied psychology*, 4(1):25–29.

Thurecht, R. L., Pelly, F. E., and Cooper, S. L. (2018). Dietitians' perceptions of the healthiness of packaged food. *Appetite*, 120:302–309.

Thurecht, R. L., Pelly, F. E., and Cooper, S. L. (2020). The influence of current food and nutrition trends on dietitians' perceptions of the healthiness of packaged food. *Public health nutrition*, 23(12):2124–2131.

Todd, P. M. and Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and brain sciences*, 23(5):727–741.

Trudel, R., Murray, K. B., Kim, S., and Chen, S. (2015). The impact of traffic light color-coding on food health perceptions and choice. *Journal of Experimental Psychology: Applied*, 21(3):255.

Turnwald, B. P., Anderson, K. G., Markus, H. R., and Crum, A. J. (2022). Nutritional analysis of foods and beverages posted in social media accounts of highly followed celebrities. *JAMA Network Open*, 5(1):e2143087–e2143087.

Turnwald, B. P., Boles, D. Z., and Crum, A. J. (2017). Association between indulgent descriptions and vegetable consumption: Twisted carrots and dynamite beets. *JAMA internal medicine*, 177(8):1216–1218.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

U.S. Department of Agriculture (2019). Fooddata central.

US Food And Drug Administration (2016). Use of the term 'healthy' in the labeling of human food products: guidance for industry.

van der Bend, D. L. and Lissner, L. (2019). Differences and similarities between front-of-pack nutrition labels in europe: A comparison of functional and visual aspects. *Nutrients*, 11(3):626.

Van Der Merwe, D., Kempen, E. L., Breedt, S., and De Beer, H. (2010). Food choice: student consumers' decision-making process regarding food products with limited label information. *International Journal of Consumer Studies*, 34(1):11–18.

Van Kleef, E., Van Trijp, H., Paeps, F., and Fernandez-Celemin, L. (2008). Consumer preferences for front-of-pack calories labelling. *Public health nutrition*, 11(2):203–213.

vanDellen, M. R., Isherwood, J. C., and Delose, J. E. (2016). How do people define moderation? *Appetite*, 101:156–162.

Vanderlee, L., Franco-Arellano, B., Ahmed, M., Oh, A., Lou, W., and L'Abbé, M. R. (2021). The efficacy of 'high in'warning labels, health star and traffic light front-of-package labelling: An online randomised control trial. *Public Health Nutrition*, 24(1):62–74.

Vecchio, R. and Cavallo, C. (2019). Increasing healthy food choices through nudges: A systematic review. *Food Quality and Preference*, 78:103714.

Vidal, L., Ares, G., Machín, L., and Jaeger, S. R. (2015). Using twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations". *Food Quality and Preference*, 45:58–69.

Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207.

Weber, E. U. and Johnson, E. J. (2009). Mindful judgment and decision making. *Annual review of psychology*, 60:53–85.

Weitkamp, E. and Eidsvaag, T. (2014). Agenda building in media coverage of food research: Superfoods coverage in uk national newspapers. *Journalism Practice*, 8(6):871–886.

Werle, C. O., Trendel, O., and Ardito, G. (2013). Unhealthy food is not tastier for everybody: The "healthy= tasty" french intuition. *Food Quality and Preference*, 28(1):116–121.

Whalen, R., Harrold, J., Child, S., Halford, J., and 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.

World Health Organisation (2021). Obesity and overweight.

World Health Organization (2021). Implementing policies to restrict food marketing: a review of contextual factors. *World Health Organization*.

Wyse, R., Jackson, J. K., Delaney, T., Grady, A., Stacey, F., Wolfenden, L., Barnes, C., McLaughlin, M., and Yoong, S. L. (2021). The effectiveness of interventions delivered using digital food environments to encourage healthy food choices: A systematic review and meta-analysis. *Nutrients*, 13(7):2255.

Yao, Z., Sun, Y., Ding, W., Rao, N., and Xiong, H. (2018). Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the eleventh acm international conference on web search and data mining*, pages 673–681.

Yates, A. A. (2006). Which dietary reference intake is best suited to serve as the basis for nutrition labeling for daily values? *The Journal of nutrition*, 136(10):2457–2462.

Zamani, M., Schwartz, H. A., Eichstaedt, J., Guntuku, S. C., Virinchipuram Ganesan, A., Clouston, S., and Giorgi, S. (2020). Understanding Weekly COVID-19 Concerns through Dynamic Content-Specific LDA Topic Modeling. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 193–198. Association for Computational Linguistics.

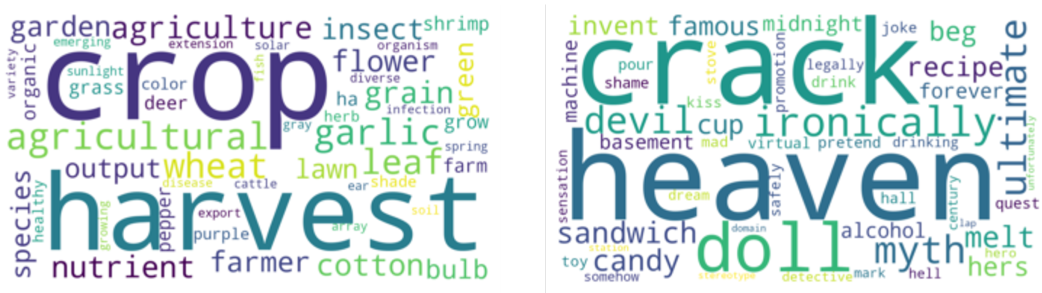
Zou, W. and Bhatia, S. (2021). Judgment errors in naturalistic numerical estimation. *Cognition*, 211:104647.

# Appendix A

# All Word Clouds

**Figure A.1**

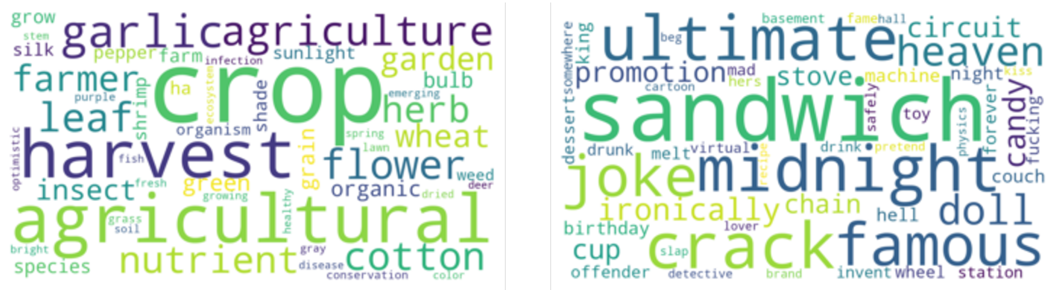
Study 1B: Healthiest and Unhealthiest “Other” Words




*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the registered dietitians in Study 1B.

**Figure A.2**

Study 1C: Healthiest and Unhealthiest “Other” Words




*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names and images in Study 1C.

**Figure A.3**

Study 2A Control Condition: Healthiest and Unhealthiest “Other” Words


*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names in the control group of Study 2A.

**Figure A.4**

Study 2A Calorie Labeling Condition: Healthiest and Unhealthiest “Other” Words


*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names and calorie labeling in the experimental group of Study 2A.

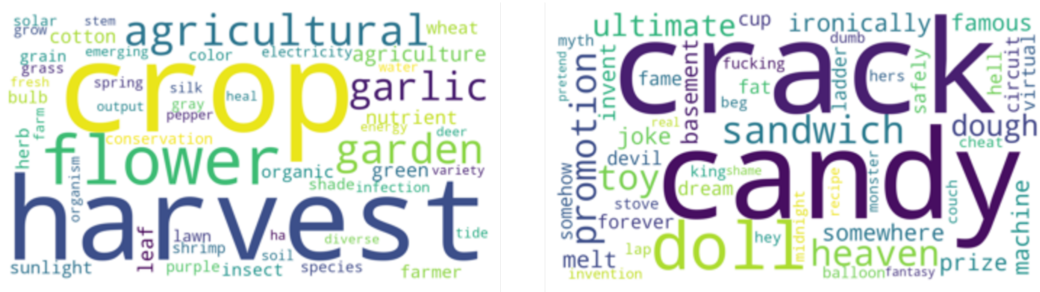
**Figure A.5**

Study 2B Control Condition: Healthiest and Unhealthiest “Other” Words


*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names in the control group of Study 2B.

**Figure A.6**

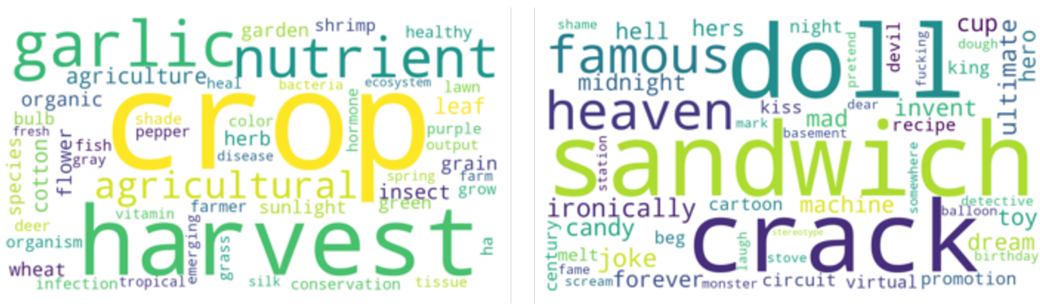
Study 2B Front-of-Pack Labeling Condition: Healthiest and Unhealthiest "Other" Words




*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names and front-of-pack labeling in the experimental group of Study 2B.

**Figure A.7**

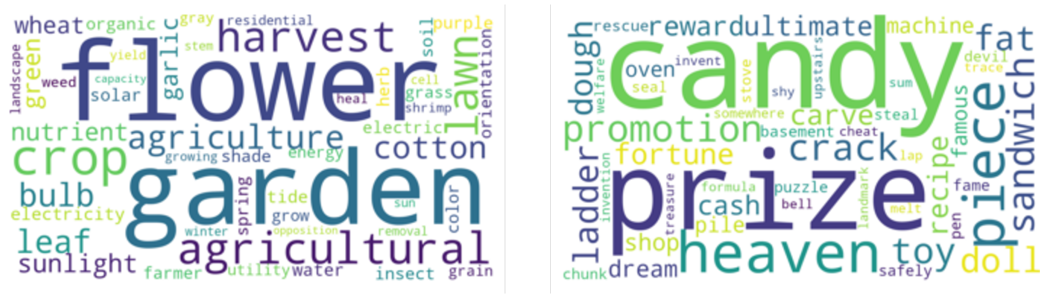
Study 2C Control Condition: Healthiest and Unhealthiest "Other" Words




*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names in the control group of Study 2C.

**Figure A.8**

Study 2C Traffic Light Labeling Condition: Healthiest and Unhealthiest "Other" Words




*Note.* Words with the highest (left) and lowest (right) predicted healthiness ratings based on the Vector Representation Model estimated for the lay people who were presented with food names and Traffic Light Labeling in the experimental group of Study 2C.