Title
Investigating Preferred Food Description Practices in Digital Food Journaling

ISBN
9781450384766

Authors
Silva, Lucas M
Epstein, Daniel A

Publication Date
2021-06-28

DOI
10.1145/3461778.3462145

Peer reviewed

Investigating Preferred Food Description Practices in Digital Food Journaling

Lucas M. Silva
University of California, Irvine
silvald@uci.edu

Daniel A. Epstein
University of California, Irvine
epstein@ics.uci.edu

ABSTRACT

Journaling of consumed foods through digital devices is a popular self-tracking strategy for weight loss and eating mindfulness. Research has explored modalities, like photos and open-ended text and voice descriptions, to make journaling less burdensome and more descriptive than traditional barcode and database searches. However, less is known about how people prefer to journal foods when less constrained by limitations of databases, natural language processing, and image recognition. We deployed a food journal prototype supporting varied devices and input modalities, which 15 participants used to journal 1008 food logs over two weeks. Participants had diverse strategies for indicating what and how much they ate, varying from ambiguous foods to specifying varieties and using different measurements for clarifying amount. Some strategies were interpretable by natural language food identification and image classification services, while others point to open research questions. We finally discuss opportunities for accounting for variance in food journaling.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **HCI design and evaluation methods** → User studies;

KEYWORDS

Food journaling, prototype, personal informatics, multimodality

ACM Reference Format:
Lucas M. Silva and Daniel A. Epstein. 2021. Investigating Preferred Food Description Practices in Digital Food Journaling. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28–July 02, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3461778.3462145

1 INTRODUCTION

Food tracking, or digital food journaling, has become one of the most popular self-tracking strategies to help with self-awareness and eating mindfulness, with 42% of U.S. adults having tried an app for diet or nutrition tracking as of 2017 [10, 30, 35, 37, 81]. Tracking of food intake has been shown to help people achieve health goals such as losing weight [18, 67, 81] and managing chronic diseases (e.g., diabetes) [27, 32, 48, 63], identifying intolerances [53, 75], and making healthier food choices [58]. Commercial applications typically support people in journaling the foods they eat through lookups to food databases and scanning barcodes on packaged foods, enabling them to aggregate a record of the foods they eat and monitor how their daily intake aligns with calorie or nutrient goals [6, 52, 77, 81].

Although food tracking can promote health benefits, it is widely regarded as burdensome, requiring a person to reliably journal to produce useful logs and contend with challenges around journaling accurately [30, 31, 52]. For example, some foods may not appear in food databases (e.g., foods from cultures the database was not designed to support), and some social contexts may make journaling uncomfortable or awkward [31]. In light of food tracking challenges, various research efforts have sought to minimize tracking burden by examining input modalities (i.e., methods of input) beyond barcode and database searches, such as photos [30, 67] and voice memos [62, 77]. Several efforts have also invested in automating or complementing food tracking through eating moment detection [11, 12, 26, 89], food image analysis [13, 45, 65, 83], and identifying food consumption in natural language descriptions (e.g., social media posts) [2, 25, 79]. Other devices pose further opportunity for people to track their foods as they navigate different contexts throughout their life, such as through conversational interactions with increasingly-available voice assistants (VA) [38, 68, 76].

Food journaling can be used to support open-ended awareness and mindfulness goals [10, 37] as well as calorie and nutrient-consumption goals [6, 52, 77, 81]. To support awareness and mindfulness, technology has often leveraged flexible food journaling through text descriptions or photos to allow people to self-describe their food or eating moments however they desire [10, 30]. Towards nutrient consumption goals, substantial work has examined how to recognize the foods and amount eaten to convert these input modalities into logs which contain consumed nutrients. For example, research in computer vision [16, 45, 82] and crowdsourcing [67] has examined labeling foods in images, while work in natural language processing [55, 56] has sought to identify what and how much a person ate from a text description. Research has also looked at improving the coverage of food databases [52].

As food journaling becomes incorporated into more devices and systems, people will have access to increasingly varied methods of tracking their foods in their daily routines. However, less is known about how people wish to record their foods when under fewer technology constraints around recognition, accurate entry, and desire for recall. For example, input modalities can include varying levels of detail about the foods people eat, from listing ingredients to describing a high-level category food fall under. Additionally, people often incorporate contextual information of where they ate and who they ate with into flexible logs [30]. Understanding how

DIS '21, June 28–July 02, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8476-6/21/06.
https://doi.org/10.1145/3461778.3462145

people choose to describe what they eat in real-world settings can offer suggestions for how technology can better support people's preferred ways of journaling, such as opportunities for technology to assist in clarifying descriptions, adding context when desired, suggesting areas where recognition can improve, or accounting for variance in how people describe their foods.

To understand how people prefer to journal their foods when under fewer constraints, we developed and deployed ModEat, a lightweight prototype for journaling on phones, computers, and voice assistants, that supported different input modalities, like barcode scanning, free text entry, voice logging, photo-taking, and simulating a database search. 15 participants journaled 1008 food logs with the prototype over two weeks. In analyzing food logs and post-deployment interviews, we identify preferences and strategies for describing foods. To understand gaps between description and recognition, we further investigate how participant's logs were and were not interpretable by commercial natural language processing (NLP) and food image classification services. Through these analyses, we note high variance in how people prefer to describe what they eat, both between individuals based on goals and among individuals based on their foods and circumstances. We contribute:

- An understanding of how people choose to describe what and how much they ate when less constrained by recognition limitations, through the deployment of ModEat, a flexible technology prototype. Participant's journal entries varied in how they described the granularity, specificity, amount, and context of the foods they ate. Participant's descriptions were still typically interpretable by NLP and Computer Vision services, but were less effective at evaluating more ambiguous descriptions or unclear food packages.
- An understanding of how people's food journaling goals and modality use influence their preferred journaling strategies. Input modality tended to influence the level of granularity and specificity participants used to describe foods, aggregating multiple foods into a single input more often and describing foods less specifically in more flexible input modalities, like plain text and voice descriptions. Participants interested in quantifying their nutrition typically included formal measurements or counts of distinct items they ate, while people with less quantitative focus tended to not indicate how much they ate.
- Design recommendations for addressing and accounting for variability in how people prefer to describe foods. We suggest that designs could help mitigate variance in journal entries through conversational approaches, but can also acknowledge or leverage ambiguity to promote reminiscence. By combining with general-purpose classifiers, recognition services could also detect more contextual information in people's text descriptions and images.

2 BACKGROUND

Our work builds on previous self-tracking research examining approaches for manual and automated food journaling, leveraging prototype deployment for eliciting people's perspective on use of technology in real-world settings.

2.1 Food Journaling

Self-tracking technology, or personal informatics technology, aims to help people monitor and understand their habits [57]. Food journaling is one of the most popular self-tracking domains, helping people monitor and understand their food related practices [39] and change their behaviors towards healthier eating habits [50]. While food journaling can be done on paper, it has been supported in technology through barcode scans of packaged foods [77] and database lookups [6, 52, 77, 81]. These strategies aim to accurately identify nutrient information from food databases and barcode libraries. These techniques are pervasive in commercial apps such as MyFitnessPal [66], WW (formerly Weight Watchers) [86], and Lose It! [59], as well as various research that employ food journaling systems [6, 33, 52, 81]. However, people often find needing to search for and correctly identify every food eaten in food databases tedious [52], with unreliable information or difficult to find specific foods and amounts [31]. Barcode scanning can lower this burden, but can potentially nudge people from eating fresh and healthier foods in favor of packaged ones [31]. Recent work has also tried to lower the burden of database searches. In the design of EaT [52], Jung et al. leverage a search-accelerator for narrowing search results in a large food database. Participants found the search-accelerator easy to use and effective for reducing typing, but logged accuracy was impaired when users did not provide details for some composite foods.

In general, manual self-tracking requires substantial effort [23], with food tracking in particular introducing additional challenges that make the practice burdensome. Cordeiro et al. identified several barriers to typical database and barcode food journals that negatively impact food tracking practices [31]. People often do not want to journal in social situations because of a perceived stigma, forget to journal and fall out of the habit, find homemade or ethnic foods more difficult to journal, struggle to contend with unreliable food databases, and feel shame or judgment when not reaching a food-related goal due to prominent calorie and nutrient goal features [31].

Research has examined approaches to make manual food journaling more flexible to promote eating mindfulness and awareness rather than collecting calorie or nutrient metrics. Photos, free text, and voice inputs have been found to be feasible ways of recording and describing foods [15, 27, 30, 37]. Some commercial apps, such as YouFood [88] and Ate [9], have also leveraged voice and photo-based journaling. Free text input has typically been incorporated to complement photo entries [27, 30, 37], often to add more details to assist social contacts [28, 60] or clinicians [27, 34, 61] in interpretation.

Various research efforts have further examined automation for lowering food journaling burden. These efforts have been examined detecting eating moments through sound (e.g., chewing noises) [5, 11, 71], automatic photo-taking with wearable cameras [12, 80], movement or proximity with necklace-like wearables [26, 89], and combining multiple sensor modalities [70]. Research has also sought to estimate food volume and label identified foods through crowdsourcing nutritional estimates [67], computer vision [13, 65, 82, 83, 85], and natural language processing (NLP) [55, 56].

With increased power of computer vision and machine learning, food recognition from images have been introduced into systems for ingredient identification and nutritional estimation. Im2Calories uses CNN-based deep learning to identify foods from restaurants, interpreting their food volume and their nutritional values [65]. Similarly, Menu-Match logs calorie intake estimated from image classification and food databases, but also has a focus on restaurant foods [13]. Ingredients or nutrients recognized can further be used to create more abstract representations, such as digital postcards summarizing what was eaten for later reflection [79] or reverse-engineering recipes from photos [19, 74]. Different from using deep-learning techniques, Yang et al. proposed representing foods as pairwise statistics over image pixels, reasoning that this can indicate spatial combinations of ingredients (e.g., a "bread" pixel next to a "cheese" pixel) [87].

Other systems have proposed the use of NLP for automatically interpreting foods from spoken or open-ended written descriptions. Korpusik et al. proposed the use of deep learning for semantic mapping foods in text based meal descriptions with searches in USDA's [40] food database [55]. In a different contribution, Korpusik et al. incorporated this technique onto a food journaling app, using both text-based descriptions and voice input with speech recognition for then NLP and database mapping [56]. Studies have also aimed to identify and classify foods in free text descriptions of foods posted to social media to understand community-level eating practices [2, 25].

Although these systems posit that automation can lower food journaling burden, complete automation can be inaccurate and is still far from being practical [42, 58, 63]. Choe et al. further highlight that fully automating data capture can reduce awareness and engagement, suggesting that self-monitoring technology balance manual and automatic tracking [23]. We expand opportunities for supporting manual and semi-automated food journaling by understanding people's journaling preferences when less constrained by current recognition limitations.

2.2 Elicitation of In-Situ Everyday Technology Use

Elicitation studies are often used to understand people's preferences for interacting with technology, such as input methods. For example, they have been used to discover preferred gestures for interacting with mobile devices [73] or interactive surfaces [84]. Although these studies uncover ways people wish to interact with technology, they can fall short of considering real-world contexts. For example, people's preferences might be influenced by their social and environmental contexts, or additional factors unaccounted for in studies that are not "in-the-wild" [49]. In self-tracking, studies have used different approaches to circumvent constraints of in-lab elicitation studies. For example, Gorm et al. suggest that participant-driven photo taking can elicit in-situ technology use, such as for understanding activity tracking practices [43]. Gouveia et al. similarly leveraged video recordings from wearable cameras to understand people's use of activity trackers in daily life [44]. Other self-tracking studies have instrumented the deployed apps to understand their use, for example to identify how often participants engaged with a particular feature [44, 64] or evaluate novel interactions and approaches [46, 52, 60]. Similarly, we deploy a flexible prototype, ModEat, to understand people's preferred interactions with technology in real-world settings.

3 METHODS

We created and deployed a prototype to understand how participants would like to journal their food when less constrained by recognition or a desire for recall.

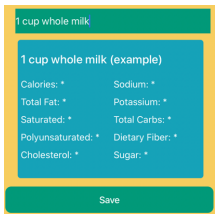
3.1 The ModEat Prototype

We developed ModEat (Figure 1), a multiplatform system available for Android and iOS phones, computers, and Amazon Alexa and Google Assistant voice assistants. We aimed to include common input techniques for food journaling in ModEat that prior research has suggested to support calorie and nutrient-consumption goals or open-ended awareness and mindfulness goals (e.g., [30, 66, 77, 81], see section 2.1), while supporting many of the devices people frequently interact with. ModEat for phone supported six different input modalities: text, voice log, picture from device camera, simulated database search, website URL, and barcode scanning. ModEat for computer ran on web browsers supporting the same input modalities as the phone, with images supported as uploads and barcodes are typable. ModEat for VAs supported conversational interaction commands, allowing people to create a new food description (e.g., "*journal green eggs and ham*") or request and hear their previous journal entry ("*read last entry*"). ModEat for computer and mobile supported reviewing previous entries by displaying the result (e.g., the text description, the numeric barcode, a simulated database search), with voice logs being displayed as text. We also implemented ModEat for Apple Watch via voice logging, but none of the participants regularly wore an Apple Watch, so we do not report further on watch entry.

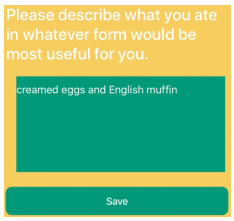
We sought to avoid incorporating suggestions for what or how to journal foods in ModEat. We intentionally did not add food recognition features to ModEat (e.g., image recognition, database searches, look up UPC barcodes), instead asking participants to suspend belief about feedback and journal as if receiving idealized responses. Doing so spared participants from limitations of current technology (e.g., incorrect or missing foods in databases, images which could not be recognized, barcodes which could not be identified), while prompting consideration of preferred entry methods.

3.2 Participants

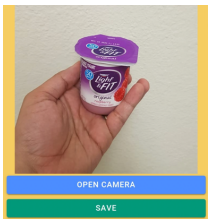
Our study was approved by our university's IRB prior to recruitment. We advertised our study through a screener survey sent to local mailing lists and subreddits related to food tracking or cities close to our university. We targeted the recruitment of people with prior journaling experience or interested in starting to journal. We primarily recruited participants with prior experience to ensure participants were highly-motivated to pursue a journaling goal, and would therefore carefully consider how their descriptions of foods would support their goals. Participant's past experiences also made them aware of the capabilities and recognition constraints of traditional food journaling approaches, enabling them to enact different uses of ModEat if they wished and allow them to journal



(a) Simulated database search

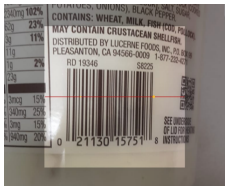


(b) Text description input

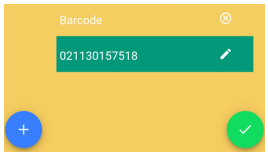


(c) Photo Input

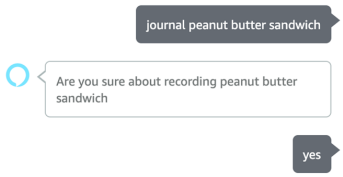
(d) URL input



(e) Scanning barcode of a packaged food



(f) Barcode input after scanning



(g) Visual representation of voice input with Alexa

Figure 1: Examples of modality inputs on ModEat phone and VA.

Table 1: Nearly all participants had prior experience using digital food journals, but had a mix of awareness and quantitative goals.

ID	Gender	Occupation	Age	Journaling Experience	How Journaled	Journaling Goal
P1	Female	Designer	36	4 years	Calendar	Awareness
P2	Female	Massage Therapist	35	2.5 years	Paper, LoseIt, MyFitnessPal	Quantitative
P3	Male	Civil Engineer	33	2.5 months	Spreadsheet	Awareness
P4	Male	Engineering Manager	38	1 month	Paper	Awareness
P5	Female	Student	28	3 years	MyFitnessPal, Self-made app	Quantitative
P6	Female	Student	25	~10 months	Cronometer	Quantitative
P7	Female	Retail	30	3 months	Paper	Awareness
P8	Female	Accounting Clerk	27	1 month	Spreadsheet	Awareness
P9	Male	Engineer	31	-	-	Awareness
P10	Male	Student	28	2 years	MyFitnessPal	Quantitative
P11	Female	Researcher	50	~2 months	FitDay	Awareness
P12	Male	Engineer	43	"On and off"	MyFitnessPal	Quantitative
P13	Female	Academic Librarian	44	2 years	MyFitnessPal	Awareness
P14	Woman	Student	33	3 years	MyFitnessPal	Quantitative
P15	Male	Drafting Design	31	2 months	MyFitnessPal	Quantitative

with more freedom to go beyond current recognition constraints. In addition, we required participants speak English and be 18 years or older. We also required that participants have access to a phone, computer and Amazon Alexa or Google Assistant. In case a participant did not own a VA, we offered to lend one if they were located near our university. Participants used ModEat for two weeks between February and April 2020. Participants were offered $30 as compensation for full participation.

55 people responded to our screener survey, out of which 33 satisfied our requirements and 18 responded to our contact. One participant dropped out of the study due to family health issues and two were dropped because they became unresponsive during deployment. The remaining 15 participants had a median age of 33 (range: 25 – 50), 8 identified as female, 6 as male, and 1 as a woman. While P9 had no prior food journaling experience, the other participants had a median journaling experience of 10 months (range: 1 month - 4 years). Participants had various personal motivations for journaling that could fall under one of two categories: quantitative goals, that focused on various numerical information (e.g., calories, micro or macro nutrients); or awareness goals, that focused on broad food consumption information (e.g., eating more greens, frequency of snacking, general healthy eating). Table 1 summarizes participant related information.

3.3 Study Procedures

Each participant went through an initial 30-minute onboarding interview to understand the study's goal, learn about the ModEat, configure it on their devices, and tell us about their journaling

experience and goals. For a few participants (P5, P6, P7, P19) this interview was in-person in public locations, but an increase in the spread of COVID-19 led every other participant-researcher interaction to be conducted through remote video calls. We encouraged participants to journal their foods according to their personal food goals and with whatever input modality they preferred at a given moment, thinking past the recognition constraints of similar current technology they had used or seen before. Participants were also provided with a manual describing ModEat's configuration and features for later reference if needed (included in the supplemental materials).

Participants used ModEat for their daily food tracking for at least two weeks (mean 14.7 days, min 14, max 16). Other food journaling studies have similarly deployed systems between 2 to 4 weeks to understand participant's regular use [37, 51, 81, 89, 90]. To help the research team understand the circumstances surrounding each journal entry, participants also answered a daily survey questionnaire. The questionnaire showed participants the entries they made that day, asking them to describe where they ate (e.g., home, restaurant), whether they ate with others, the type of meal (e.g., full meal, snack), and when they journaled relative to when they ate (e.g., long before, long after, while eating).

We conducted a one-hour post-deployment interview with each participant to discuss their experiences with ModEat. During the interview, we showed participants their food logs and asked questions to clarify how and why they chose their logging techniques. We also asked participants about the limitations they experienced with current journaling strategies and what they would ideally wish to be able to do with food journaling technologies. Participants were compensated at the end of the interview and returned any lent device in a socially distant manner (e.g., leaving device in the porch).

3.4 Data Analysis

All journal entries were separated by input (e.g., text description, barcode, photo), resulting in 1008 individual inputs. The first author followed thematic analysis [17] to analyze the food logs, first open coding and then discussing themes with the research team. After refining definitions and coding criteria, the final codebook contained 39 codes in 12 categories, such as how many food items were present in a log, how specifically foods were described, and if and how logs described food amounts. For example, the code category *amount* had the subcodes *numeric scale*, *numeric only*, *broad*, *comparative/reference*, *non-standard*, and *non-quantified*). Two authors independently coded the same 10% of inputs, reaching near-perfect agreement on 34 codes (Cohen's $\kappa \geq 0.8$) and substantial agreement on the remaining 5 codes (Cohen's $\kappa \geq 0.6$). They discussed and resolved differences in code application, then the first author coded the remaining data. All final interviews were audio-recorded and transcribed through a university-approved vendor. Authors reviewed interview transcripts to understand participant's reasons for how they described their foods to support the analysis of the food logs.

Once all logs were coded, we used logistic regression to quantitatively analyze the influence of contextual factors surrounding the creation of journal entries based on the frequency of specific codes (e.g., whether journal entries created with certain modalities were more likely to be more granular or describe amounts). We examined four contextual factors from the daily surveys as fixed effects: modality used, how many others they ate with, meal type, whether they journaled before or after eating. We treated participant id as a random effect to account for individual differences in how participants described foods (e.g., participants who were more likely to include amounts). We corrected for multiple comparisons in post-hoc tests between levels of fixed effects with Tukey corrections.

To understand how participant's preferred methods of food journaling aligned with traditional approaches to food recognition, we ran participant's logs through commercially-available recognition services. It is not our intention to evaluate or compare the overall accuracy or quality of these methods for food identification, rather to explore how they might need to adjust to people's strategies for describing their food.

We submitted database search, text, and voice input descriptions to commercially-available NLP services. We are not aware of prior research comparing and ranking the quality of NLP services for foods, so we ran three services (Nutritionix [69], Edamam [36], and Spoonacular [78]) regarded as highly used and frequently mentioned in top lists (e.g., [72]) on a random 10% of journal inputs, comparing the results against our manual inspection of the descriptions. These services accept requests with food descriptions in text and aim to return a list of described foods and their amounts, including an amount unit (e.g., grams, cups), calories, and nutrients (e.g., fats, sugar). For example, a text input of "*100g rice and 2 eggs*" should identify "*rice*" and "*eggs*", and "*100 grams*" and "*2*" as their respective amounts, returning calorie and nutritional information. This differs from non-NLP services that only do direct database searches for provided food items, such as the USDA FoodData Central [40]. The NLP services proved to have different levels of success for a set of 10% random journal inputs. All three services identified amounts fairly similarly (37.6%, 33.6%, 35.6% of inputs), but Nutritionix was more successful in identifying all food items in an input (78.2%), versus Edamam (37.6%) and Spoonacular (57.43%). Furthermore, Nutritionix failed to identify any food item or amount in just 1.9% of inputs, while Edamam in 21.7% and Spoonacular in 19.8%. Therefore, we chose to analyze the remaining inputs with results from Nutritionix.

We also used commercially-available image classification services to understand opportunities for improving recognition of the pictures people used to represent their foods. Similar to NLP services, we chose three popular services regarded in top service lists (e.g., [1]): Clarifai [29], Google CloudVision [41], and Amazon Rekognition [4]. Clarifai provides a service module specific for classification of food images, while the other two are publicized as image classification more generally and offer food detection as an example. These services return a list of identified elements paired with a confidence level (e.g., a percentage from 0-100, probability from 0-1). For example, a query for an image of a chocolate cake could result in "*[chocolate(1.00) cake(1.00) brownie(0.94) ... flour(0.38) pumpernickel(0.34) cookie(0.32)]*". We ran the three services over 54 of our full set of 60 images, again comparing against our manual inspection of the foods present in the images and discarding 6 images where we could not manually identify any of the foods. We took a

conservative approach to recognition, considering all foods identified by the service with confidences above 0.8 as recognized. For instance, we coded an image as successfully identified if the classifier service identified any element present in an image of a ham and cheese sandwich (e.g., "*bread*", "*cheese*") with at least 0.8 probability, regardless of whether foods which were not present were identified with high confidence (e.g., "*cookie*" at 0.92 confidence).

We also searched barcode inputs in a barcode search website [8], searching for unidentified items using the barcode search in MyFitnessPal [66], which also leverages user-created database items.

We quote participants with PXX when presenting quotes from interviews or journal descriptions they created.

3.5 Limitations

We acknowledge our limited sample size, and although relatively diverse in gender and occupation, findings might not be generalizable to the journaling practices of different age or cultural groups. For instance, low-income communities have particular needs and expectations around food journaling technologies [46], and care needs to be taken in applying our findings to these cultural settings. Furthermore, while our participants' high degree of interest and prior experience using food journaling technologies led them to carefully consider how their approaches to journaling in ModEat would support their goals, their prior experience may have also influenced a few participants towards approaches they were familiar with. Though most participants used and appreciated a range of input techniques, a few participants described intending to leverage database searches in ModEat in similar ways to their journaling prior experiences. None of our participants had a disease diagnosis or management goal, although it is a commonly-studied motivation for food journaling [48, 53, 75, 90]. We suspect that people with such a motivation might be inclined to emphasize specificity in their food descriptions and may have other different journaling preferences from our participants.

Study deployment coincided with the COVID-19 pandemic, with 10 participants (P1-5, P7, P9, P11, P12, P14) mentioning feeling significant impact on the eating habits and stress levels. For instance, participants that would frequently eat at work or at restaurants almost exclusively ate at home during the study. The pandemic also influenced available foods, such as P13 that mentioned "*I would love to eat more fresh foods . . . [but] I can't go to the grocery store multiple times*". Other than the emotional and social consequences, only P9 and P13 felt their food description practices were impacted by the pandemic, with P9 reporting being "*a little more observant*" to detailing food compositions and P13 doing the opposite by relaxing her diet restrictions. Overall, participants still used regularly ModEat to explore varied journaling strategies, despite the pandemic influencing what they ate and changing contexts surrounding how they ate.

4 FINDINGS

Participants used the ModEat prototype to make 659 food journal entries with 1008 modality inputs (average 1.53 modality inputs per entry, max 13). Participants journaled fairly frequently, averaging 2.98 entries per day (min 1.07, max 4.26). Out of the three descriptive inputs (database search, text description, and voice description), database search was the most used (37.8%), followed by text description (27.4%) and voice description (23.1%). Images and barcodes were used 60 (5.9%) and 51 (5.1%) times, while URLs were recorded 7 times (0.7%).

Overall, participants tended to use database search, text, and voice descriptions in similar ways, often describing their food choices and amounts. Participants varied in how they used those modalities to describe and measure food. Descriptions varied in granularity and specificity, occasionally captured contextual information, and indicated amounts using measurement scales or numeric values alongside subjective measures, but entries were occasionally ambiguous or unclear. Similarly, participants varied in how they used images to depict their food, such as arranging foods for aesthetics and clear amount compositions, use of stock images, and packages. This input variability had consequences for the recognition and performance of commercially-available NLP and image classification ML models, with some styles of entry more accurately interpreted than others.

4.1 Granularity

We define the *granularity* of a food log as the quantity of food items present in a single input log. We observed that food descriptions (text input, database search, voice input) were either single food item, a single item decomposed into its requisite ingredients, or aggregated foods. As described in Table 2, most food entries were composed of a single item (62.9% e.g., "*1 cup blueberry*", "*fajitas*"). However, participants occasionally described single foods with detailed ingredient compositions, such as the ingredients in a sandwich or a salad (8.2% of inputs). Some of these inputs had a food's common name followed by its composition (e.g., "*breakfast burrito with a whole wheat tortilla, two eggs, bacon. . .*", P14), while others described the ingredients without indicating a common name (e.g., "*2 tortilla with butter and honey*", P6).

Participants also regularly aggregated distinct food items into a single input (28.9% of descriptive modalities), averaging 3.09 foods per input when they aggregated (min 2, max 9). These differ from decomposed single food's in that they typically combined foods eaten together in a single event (e.g., meal, a snack), but represented distinct foods. We classified 101 aggregated food inputs (39%) as main course dishes journaled with one or more side dishes. For example, P13 logged "*egg omelet and side salad*", P9 logged "*wonton soup, Chinese book choy, rice, breaded shrimp. Orange*", P4 logged "*Veggie enchiladas, half a cookie, grapes, pistacios [sic]*". 60 aggregated food inputs (23.3%) were foods alongside drinks, such as "*coffee and banana*" (P5) and "*tea and chips*" (P9). Participants tended to aggregate entries more often when eating with others versus alone ($Z=2.04$, $p<0.05$, 95% CI 2%-90% more likely to aggregate), perhaps suggesting that participants tended to aggregate when in social situations where they wanted to journal multiple items quickly.

Input modality tended to influence the granularity with which participants entered food ($\chi^2(2, N=890)=89.56$, $p<0.001$), but we did not observe a statistically significant impact of food journaling goals on granularity ($p=0.15$). Figure 2a shows that nearly all database searches were of single food items (94.5%), versus about half of voice input entries (52%), and a quarter of text inputs (28.6%). P6 described

Table 2: Examples of granularity and specificity from participant food journal entries.

	Generic	Specific	Varietal	# of Inputs
Single Food	"*Pizza*"	"*Chicken*"	"*4 oz chicken thigh*"	560 (62.9%)
	"*tea*"	"*Milk*"	"*240 ml whole milk*"	
	"*mixed vegetables*"	"*Orange*"	"*Hawaiian beef*"	
Decomposed Single Food	"*Random greens in tortilla*"	"*Hummus and cheese sandwich*"	"*protein shake with almond milk 1 scoop protein powder*"	
	"*vegetable soup*"	"*cauliflower rice 2 servings*"	"*Trader Joe's, tahini, pepita, & apricot slaw kit*"	73 (8.2%)
	"*chips soup*"	"*a bowl of rice and 3 meatballs*"	"*drunken chicken noodles*"	
Aggregated Foods	"*slice of pizza with side salad*"	"*coffee and oatmeal*"	"*pizza and chicken wings*"	257(28.9%)
	"*taco and burrito*"	*"spaghetti and half an orange"*	"*orange juice, egg roll biscuits"*	
	"*sandwich and steamed vegetables*"	"*peanut butter bagel and coffee*"	"*bean burrito and corn salsa salad*"	
# of Food Items	128	842	607	

using database search for single food items because she considered it as keyword input, while other modalities were more appropriate for inputting multiple items. She said, "*Database was most useful generally because it's keywords. So, a lot of the time I put 'cutie tangerine' because we have tons of those and quick keywords. 'Banana', same thing. [. . .]. [Text] Description was because I would eat several different foods at one time and didn't want to have to put a bunch of different database searches down one long entry.*" Decomposed single food descriptions were provided most often in voice input (8.5%) and text (14.5%), but rarely in database searches (3.1%). Aggregating foods was prominent in text (56.9%) and voice (39.5%) inputs, but infrequent in database searches (2.4%). P9 explained that he found it easier to journal in this way *"when I want to record a handful of items, because I could just rattle off a bunch of things I ate [to voice assistant] [. . .] [or] I easily put down multiple items at the same time [in text description]*". Participant's logs varied in granularity both as a group and individually, other than P14 and P15 that mostly journaled single foods and in database searches, as shown in Figure 2b.

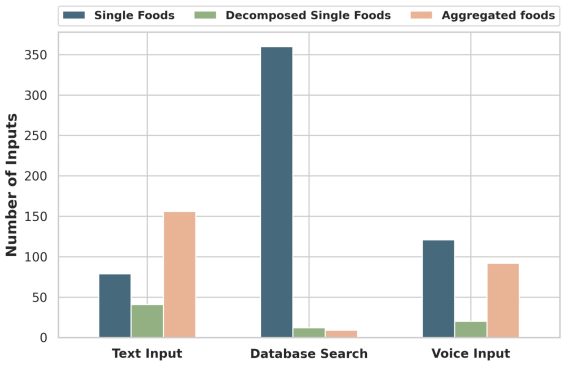
Five participants incorporated symbols to help describe or explain the components of their foods in text entry fields. For example, P6 used a "/" character as a delimiter when decomposing a food, such as "*shrimp burger / 2 shrimp patty, 1 wheat bun, 2 tsp. sriracha&mayo,* 1/4*c. spring mix*". Similarly, P7 used ":" with the same objective. P7 would also use this approach to translate and describe varieties of ethnic foods inside parenthesis: "*[Name of restaurant]: Lobster hand roll (x2), cooked scallop (hotate), albacore, salmon (sake), squid (ika), tuna (maguro), escolar (ono), yellow tail (hamachi), bluefin tuna (akami), [. . .])*". P1 used "+" to denote combinations, such as "*cheese toast with cream cheese + coffee*".

After using the ModEat prototype, some participants imagined that future journaling systems could encourage people to aggregate foods they ate in a single meal or setting through conversational approaches. Participant P15 described wanting to interact with a VA similarly to a drive-through window. He said, "*Like you drive up the window [. . .] Then I'd say, 'Alexa, journal food,' and then she just says, 'Ready.' Then I start, 'All right, I'm having, a double-double and fries and a shake.' If there's a pause, she can ask, 'Is there more?' Then I can just reply, 'Oh, add some chicken nuggets,' or whatever*". P14 had a similar idea, but felt "*it could be a burden*" to say so many details to the VA.

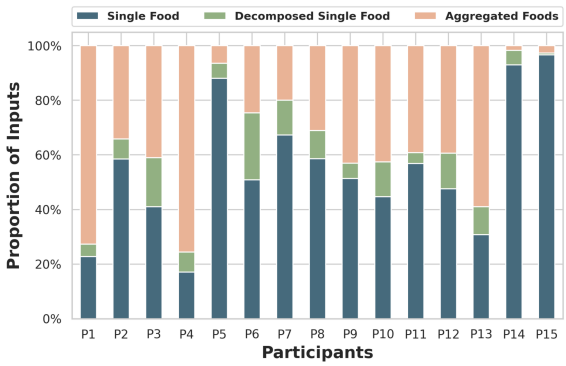
4.2 Specificity

We define *specificity* as the level of detail of a food description belonging to a particular food item, observing three levels: generic, specific, and varietal. Inputs with multiple food items could also have multiple levels of specificity (e.g., one item is specific with another being generic). A minority of food descriptions consisted of foods with generic ingredients or contents (e.g., "*dumplings. . .*" P1, "*veggie taco salad. . .*" P4), comprising 8.1% of all food items. Participants instead tended to describe foods in ways which were specific enough to distinguish between foods or ingredients of prepared foods (e.g., "*peanut butter 14 gram*" P5, "*broccoli, chicken, rice. . .*." P9; 53.4% of all food items) or further describing varietals of same food (e.g., "*red beans. . .*" P12, "*roast chicken breast. . .*", P4; 38.5% of all food items). Table 2 shows additional examples of inputs at different levels of specificity. Similar to granularity, we did not observe a statistically significant correlation between participant goal and specificity levels (p=0.83). Instead, all participants greatly varied individually in how specifically they journaled their foods, as visualized in Figure 3a.

Descriptions of aggregated foods were not always clear about how they were composed, leading to potential uncertainty or ambiguity around what was eaten. 39% of descriptive inputs with multiple food items were joined with conjunctions or prepositions like "in", "and", or "with" (e.g., "*chicken broth with rice and chicken meat*" P11), while 25% of text input and database searches with multiple food items used commas as a separation symbol (e.g., "*orange juice, egg roll biscuits*" P9), and 9.4% used both (e.g., "*coffee,* 1/2 *bagel with cream cheese*", P13). However, some aggregated or decomposed food inputs (92, or 27.8%) had unclear food descriptions, especially when lacking conjunctions or item separators. For example, descriptions like "*coffee cinnamon rolls*" (P2), "*cup of soy milk small pastries*" (P8), and "*half apple chicken link*" (P15), could

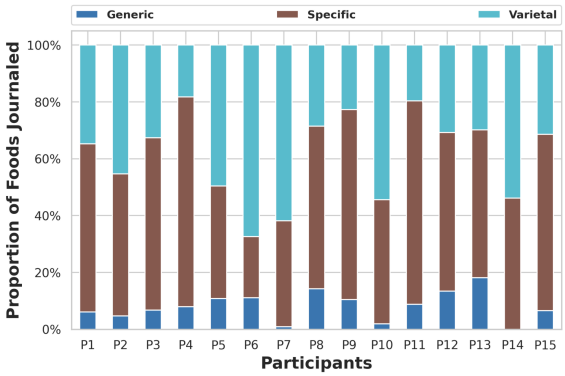


(a) Distribution of granularity per modality

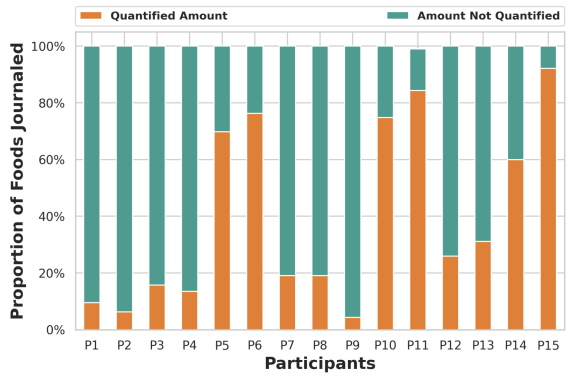


(b) Distribution of granularity per participant

Figure 2: Participants varied in the granularity they used to describe their foods in each modality, typically journaling a single item with database searches and often aggregating multiple items in a single input or detailing a food's ingredient composition with text and voice inputs. Individually, most participants entries and journaled foods varied in granularity and specificity, though a few participants consistently created single-food entries.



(a) Distribution of specificity of food items per participant



(b) Distribution of amount description per participant

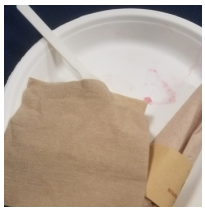
Figure 3: Specificity of food detail varied per modality while amount tended to align with personal food journaling goal.

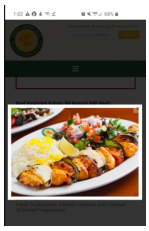
be interpreted as flavors or varieties, or separate items that were combined into a single entry. For example, P5's description of "*warrior chia bar cinnamon and apple*" could be interpreted as a bar with cinnamon and apple flavor, versus bar with cinnamon flavor and a side apple.

Database entries were more likely to include more specific items than either text descriptions or voice descriptions (Z=4.68, p<0.001, 95% CI 28%-87% more likely). Participants explained that they expected that the input to database searches would need to be specific. For example, P14 said, "*if it's lasagna, there's going to be 10,000 homemade lasagna's in the database*", expressing that the more generic description could have varied nutrient information or ingredients. P5 sought to circumvent this by decomposing foods into their elements, she said that "*you start by [searching] ingredients [. . .] So, when I use the search function for those, I would get the accurate calorie counts*". However, *"it's annoying to find all the ingredients*" (P14) and can be a "*chore to log your food, and to be accurate with it*" (P12), revealing tension between accuracy, effort and specificity for this modality.

Participants described simply not being able to add more specificity in some circumstances. P12 reflected that sometimes when eating at a restaurant, "*you can't be as finely detailed, unfortunately. [. . .] [if it is a] non-chain type restaurant, you've just got to eyeball it, there's literally no way of getting around it*". P13 had similar remarks about restaurant foods, adding that, "*If you eat out and you are logging what you had, then more than likely, all you can say is, 'I had Irish stew,' because I don't know what's in it. I don't know what oils are in it*". Some participants also felt that some foods did not warrant detailed descriptions. P14 explained that "*[my logs] will often be [with] a check-in for red wine. That is the only one I will want to use. It doesn't matter what kind of wine it is, it's more useful to me to have those carbs accounted for... The same is true for beer and any other booze*". Similarly, P5 said "*I kind of want to know that I got enough fiber, [if] I just eat enough greens*", justifying making


(a) Presence of receipt


(b) Empty plate after eating


(c) Restaurant food screenshot

(d) Package after finished eating

(e) Stock image


(f) Desktop dining


(g) Aggregated foods


(h) Separated composition

(i) Aggregated and specified food


(j) Clear food amounts

(k) Package with consumed quantity

(l) Packaged snack during eating


(m) Food being made


(n) incorrectly classified, but like (i)


(o) Leftover food journaled long after eating

Figure 4: Participants had various styles of using photos to capture foods and eating events. Participants used photos indicate amounts of foods eaten by referencing containers, leveraged stock images to represent foods similar to what was eaten, intentionally laid for better recognition of foods eaten, and took photos of foods at different stages of being eaten.

generic food logs such as "*random salad greens*" and "*random greens in tortilla*".

Burdens of entry occasionally led participants to be less specific than they wanted to be. For example, participants reported that issues with VAs not hearing or understanding them led to repetition and frustration, and also felt it led them to speak shorter descriptions in hopes of decreasing errors. For instance, P10 said, "*as you increase the amount of time using voice input, the chances for the number of errors that you would have using the voice entry increases. So, I felt compelled to use only short phrases that I could really enunciate and that I believe would be easily recognizable by the voice assistant.*" Similarly, P2 said that, "*I got the impression with it that the command had to be really short, that if there was too long of a pause it would just [finish] recording what it was*". This aligns with previous work identifying that people often shorten and simplify their sentences in the hopes of being better understood by voice interfaces [14].

When journaling with images, participants used different strategies to depict the specificity of their foods. Most images were of foods just about to be eaten, such as on plates and in wrappers (43/60). Some images were pictures of food packages (7/60), others were stock images retrieved from the web (10/60). P10 purposefully took a picture of his food alongside the receipt while eating at a restaurant to better describe what he ate (Figure 4a). In another situation, he took a picture of an empty plate (Figure 4b) alongside a text input describing its consumed content "cut fruit and 1/2 a panera berry danish". P8 would often journal with stock or menu images found online, especially for restaurant meals and when journaling long after eating (e.g., Figure 4c).

Most images had clear identifiable composition (54/60). For example, packaged food images (e.g., Figure 4d, 4e) had clear food details in text. Unclear food images typically had indistinct ingredient mixes (e.g., liquid substances in Figure 4f, 4g), an empty plate (Figure 4b), or food inside unlabeled wraps. Participants typically focused and arranged foods to be fully captured in their photos (45/50 non-stock images). In a few cases, photos modeled the food arrangement to better capture the exact amount or variety of foods eaten (4/50 non-stock images), such as Figures 4h and 4i. P7 explained that Figure 4h's arrangement was because she "*didn't want to list all the foods I was consuming. Plus, I thought the presentation*

Table 3: Examples of amount strategies grouped by granularity from participant food entries

	Scale	Numeric	Non-standard	Non-quantified
Single Food	"*soy milk 18 oz*"	"*2 cuties*"	"*small plate of seasoned almonds*"	"*spaghetti carbonara*"
	"*2 tbsp chia seeds*"	*"spaghetti squash .25"*	*"handful walnuts"*	*"granola"*
	"Cheese 1 oz"	*"4 eggs"*	*"a small glass of wine"*	*"cereal"*
Decomposed Single Food	"*PB&J / 1 low carb Mission tortilla; 1 tbsp jelly; 1 tbsp sunflower seed spread*"	"*2 tortilla with butter and honey*"	"*protein shake with almond milk 1 scoop protein powder*"	"*collard wraps with turkey*"
	"*Trader Joe's, Tahini, Pepita, & Apricot slaw kit (75g)*"	*"[a] warrior chia bar cinnamon and apple"*	*"cauliflower rice 2 servings"*	*"sesame beef tacos"*
	"*banh mi 4 oz pork 2 eggs 6 in baguette*"	*"toast with jelly"*	*"a bowl of rice and 3 meatballs"*	*"chicken noodle soup"*
Aggregated Foods	"*5oz 85% ground beef 2 tbsp sriracha 2 tbsp teriyaki; [. . .]*"	"*1/2 Calzone with salad*"	"*chips & salsa, Marbella market samples, 1 spoonful each*"	"*coffee and oatmeal*"
	"*baked potato with 1 tbsp butter*"	*"[a] baked potato with butter"*	*"medium cup vanilla flavor yogurt land with oreo"*	*"locro with rice"*
	"4 oz chicken thigh 1 tbsp Italian dressing salad mix"	*"half a sandwich in 2 oranges"*	*"one avocado one bowl of mixed vegetable and a handful beef jerky"*	*"cheese and wine"*
# of Food Items	226	331	106	907

looked nice." P6 used manual labeling on paper to discriminate the different varieties of salsa present in their meal, explaining that she arranged Figure 4i according to her lighting and that "*I tried to be pretty thorough with my recording here.*" P8 explained that for her aiming for aesthetics "*has to do a lot with what I'm used to seeing too on Instagram or social media, nice food [. . .] I would want to take a picture that would look nicer.*"

4.3 Amount

For descriptive inputs, participants used different methods to articulate how much they ate, such as using formal scales (14.3% of food items; e.g., cup, grams), numbers (21.0% of food items; e.g., "*1 roma tomato*", P15), and non-standard measures (6.7% of food items; e.g., serving, bowl, handful, slice). 10.1% of aggregated or decomposed food logs used more than one strategy for describing amounts. For instance, 33% of these mixed-amount inputs combined some food items measured using a formal scale with counted items (e.g., "*[a] baked potato with 1 tbsp butter*" P15). More than half (64.4%) of mixed-amount inputs had a quantified food item alongside food items with no amount at all, such as "*. . . plain burger, fries, 1 glass dry white wine*" (P5), "*raisin bran [cereal] and an egg*" (P4), and "*3 catfish tacos, corn tortillas, salsa*" (P3). One explanation is that some foods are more difficult to count or quantify than others, especially foods that are small, numerous, or liquids.

Participants structured their amount descriptions in various ways, reflecting however they preferred to describe their foods. 287 inputs had amount description in front of each food item (e.g., "*18 oz silk vanilla soymilk 1 oz chia seeds",* P10), while 88 indicated amount after the item (e.g., "*curried chicken sandwich .75*", P14). 2 inputs used one amount to reference every item (e.g., "*2 spoonful each of roja, fiesta, roasted*", P6). Participants occasionally used quantity descriptions in ways that they could interpret, even if not exact measures. For example, P6 acknowledged estimation by adding "~" to scale amounts (e.g., "*. . .~2 oz. salmon; ~.5c weed greens; ~.25c avocado*"). Similarly, non-standard amounts could be imprecise, referencing food containers with variable sizes such as plate, spoonful, bowl, scoop, or glass, as exemplified in Table 3

Although rare (7 inputs), participants occasionally referenced other known sizes in amounts. For example, P6 fractioned a package "*.25 pkg Trader Joe's Asian noodle salad [. . .]*" and a bottle "*Health-Ade Kombucha, pink lady apple, half bottle*". Similarly, P5 referred to a personally-known package size: "*lays baked chips subway size*". In other inputs, food amount used subjective qualifiers such as "large", "big", and "small" (16 inputs), or related to portion sizes with "slice", "cut", "entire", and "serving" (34 inputs). For example, "*one small vegetarian pizza*" (P11), "*2 slices of honey turkey breast*" (P15), "*entire pizza*" (P5), and "*bacon .4 serving*" (P14). 22 inputs used "*handful*" for snacks or ingredients. Variations included fractions (e.g., "*half handful almond*", P11), or other qualifiers ("*small handful craisins*", P15).

More than half (57.5%) of described food items had no amount clarification, and we observed no significant difference in the rate at which amounts were clarified between voice, text, or database search input modalities (p=0.36). Participant's goals typically influenced their decision for choosing whether to describe their food amounts. Typically, participants that had weight management, nutrient, or calorie-focused goals mentioned a desire for measuring their foods. For instance, P12 said, "*If you're trying to be really, really anal and accurate, you've got to remember these grams*". Likewise, P5 preferred measuring her foods, explaining that she "*wanted to make sure I get correct amount of fats logged*" when journaling "*28 gram mozzarella*" with a scale amount. In contrast, P3 was primarily interested in becoming more aware of his eating habits and explained that "*the way I would log would be more just what I ate rather than a quantity [. . .]. I would just put what I did with some qualitative things, 'I had a small plate of this', or 'I had a couple of*

this'. I wanted to make it easy to log. [...]. Just a description of the meal, rather than getting into, 'I had three eggs and 200 grams of ham and blah blah blah'." P13 similarly aimed to be "*cognizant of what they're eating*" and felt less of a need to clarify the amounts they ate. Overall, participants with quantitative goals were more likely to clarify the amount of food they ate in an entry than participants with awareness goals (Z=1.71, p<0.05, 95% CI 14%-328% more likely). Participants variated substantially in how often their food item descriptions included amounts, with 7 indicating amount in less than 25%, 3 indicating amount in more than 75%, and the remaining 5 in between. This variance is illustrated in Figure 3b.

Participants also used images to convey food quantity, usually through an angle which enabled determining volume and container dimensions. P9 felt that "*the picture would be a way to document that in terms of how many servings I had*". A sense of quantity was usually present in images that had full plates (e.g., Figure 4g), bowls (e.g., Figure 4f, 4g), cups (e.g., Figure 4j), or food units (e.g., food wrap in Figure 4a, chocolate square in Figure 4k, a banana in Figure 4j). Most images conveyed some form of food amount (57/60). However, even in cases where the foods in pictures could represent a single serving, they typically left no indication that the full amount of food shown in the image was eaten. For instance, P2 had a plate with pasta, but later revealed in a survey answer, "*I took a picture of the meal and then I forgot to eat it*".

Pictures or stock images of packages usually had visible weight descriptions on the packages, but they were sometimes still ambiguous around how much was consumed. For instance, although it could be reasonably assumed that small snack packages were fully consumed when a person journaled after they ate (e.g., the protein bar wrapper in Figure 4d), most other packages were larger (9/12 of picture and stock images). For example, P11 uploaded the same stock photo of a 2 lb. bag of oatmeal (Figure 4e) for multiple entries. Barcodes similarly varied in whether amounts could be reasonably inferred. For instance, the barcode input of a 330 ml. Vita Coco coconut water bottle (P7) could be assumed as fully consumed, whereas a barcode of a 24 oz. Kellogg's Raisin Bran cereal box (P4) is unlikely to be. In a few journal entries (5), participants clarified this ambiguity by combining inputs to detail actual eaten amounts of barcode foods. For instance, P6 made a barcode input of an 8-count tortilla package combined with a text input of "*Salmon wrap (2) ~1c. marinated salmon ~.5c microgreens [. . .]*", likely indicating that two tortillas were eaten.

4.4 Context

Participants included contextual information related to the food and eating event in a few entries (4.9%). 21 of these inputs had implicit or explicit indications of where the person ate the food. For example, P5 mentioned in a text input making a homecooked meal, "*quick homemade stir fry sauce; 15 calories*", versus another meal in a different place, "*dinner at friends house, bbq chicken with mac and cheese*". Implicit locations were present in inputs with foods from restaurant, such as "*Chinese takeout chowmein beef broccoli*" (P2), and "*WABA grill: salad, brown rice, chicken, beef, . . .*" (P3).

Like descriptive inputs, some images (14/60) also had implicit location indications. For example, participants journaled images with background elements such as desktop computers at the office or at home (e.g., Figure 4f, 4l), kitchen appliances (e.g., Figure 4m), restaurant names and time of meals on receipts (Figure 4a), and other household objects like living room tables and TV controllers. Similar to Cordeiro et al.'s findings [30], participants used images in this way to improve interpretability for reflection. However, participants seemed to focus more on capturing the nuances of their food than on capturing context. For instance, P10 said that he thought images "*sufficient for me to effectively reconstruct or eyeball how much protein I have that day*", similar to P5 that said images were useful to "*remember what you ate*". Unlike Cordeiro et al.'s [30] participants, none of ours took a picture of others present during a meal.

Participants also gave contextual information that gave more detail about when they were eating. For example, 12 inputs specified the type of meal, such as snack (e.g., "*jubes snack*", P3), dinner, dessert (e.g., "*Persian dessert, many*", P6), lunch (e.g., "*lunch: slice of pizza with side salad*", P13), and supplement (e.g., "*supplement set / 3 multi, 2 fish oil, 2 D3 [. . .]*", P6). Other inputs (35) gave glimpses at participant's routines. P3 would often tag mealtimes when journaling long after he ate, such as "*(meal eaten at 10 PM Saturday March 21st) Pasta with broccoli"*, and recurring food, such as "*same pasta, chicken, veggies, and now hot sauce*" and "*[. . .] leftover chashu and bok choy*". Similarly, P6 had a particular recurring meal that she would label as a daily mixture, detailing its varying contents each time: "*daily mix / 3 multi, 1 elderberry, 4 fiber*".

Some participants (P2, P6, P8, P12, P14) indicated that knowing meal contexts would help them reflect on their eating behaviors. For instance, P6 mentioned a desire to "*explore my emotions around my food emotions [. . .] because I'm really interested in how food would impact emotions or how my emotions impact what I eat"*, and suggested that this could be through "*writing and answering a questionnaire or photos* ". Similarly, P14 said that capturing context was lacking in her past journaling experience and could have given more insight for her food choices. She said:

> "*I have in the past thought about when I look back on my journal, on MyFitnessPal, [that] I can identify things that I felt good about eating and things that I sort of felt like, 'well that was a bit of a waste'. [I would like] Having a bit of context, if there was a way to easily visualize that somehow to sort of know, because what I would think I might find is I eat a bunch of crap, I don't need to eat late at night or during a stressful day or something like that.*"

4.5 Automatic food interpretation

Participants expressed interest in leveraging automatic interpretation of food descriptions logged. Nine participants (P2, P5, P6, P10, P11, P12, P13, P14, P15) wished that VAs could execute a background database search on described foods during the conversation or for later reflection. For instance, P11 said, "*I wish the [Alexa] VA can figure out the total calories of the food after I tell her what kind of the food I had and the quantity of the food.*". There were similar requests for interpreting text descriptions and images. P9 wished that text inputs would retrieve nutritional information, combining with database search, saying "*a blend of those two [db search and*

text input] would be great [. . .] it would give me the option of providing me the additional nutritional facts about each of the items that was in my [text] description". P9 suggested a similar feature for capturing food composition from images, comparing to Shazam, a popular music classification app: "*it would be like the Shazam of food [images]. [. . .] I think that would certainly add additional value*".

Current automated approaches were overall successful given how participants desired journaling their foods, but had some limitations depending on how participants wished to structure their entries.

4.5.1 Interpretation of natural language food descriptions. Overall, food items were generally identified correctly by the NLP systems we tested, with 80.7% of descriptive entries correctly being interpreted and returning relevant nutritional information for every food item or component (e.g., calories, micro and macro-nutrients). For example, the entry "*eggs in cheese sauce over English muffin with coffee*" (P13) was interpreted as four separate ingredients: "*eggs*", "*cheese sauce*", "*English muffin*", and "*coffee*". The remaining 19.3% inputs were not fully interpreted correctly, but 77.9% of these had at least one food item that was correctly identified. For instance, "*4 oz chicken breast 3 oz spinach 125g tamaki haiga 2 tsp soy sauce*" (P10) had all items identified except for "*tamaki haiga*". Overall, only 4.3% of inputs completely failed, either not matching any items (14 inputs) or wrongly identifying foods (24 inputs), such as "*2 tablespoons salad topper*" (P15) being classified as "salad". Six of the non-matched foods were direct references to brands, such as "*2 square 70% lindt*" (P5), while four others were ethnic foods such as, "*chapaguri*" (P7).

Modality impacted the rate at which inputs were accurately interpreted ($\chi^2(2, N=890)=36.91$, $p<0.001$). Text inputs were less likely to be interpreted correctly than voice inputs or database searches ($Z=-5.70$, $p<0.001$, 95% CI 49%-121% less likely). Text inputs had greater opportunity for at least one item not being understood due to most entries being aggregated foods, versus database searches and voice inputs that had a majority of single food inputs (Figure 3a). 79.4% of text inputs had at least one item correctly understood.

Specificity also impacted in interpretability of food descriptions ($\chi^2(2, N=890)=36.39$, $p<0.001$), with specific foods more likely to be interpreted than either generic or varietal foods ($Z=4.74$, $p<0.001$, 95% CI 32%-98% more likely). Many varietal descriptions used adjectives to describe food names, which could lead to misinterpretations and ambiguity. For instance, the voice input "*chicken eggs and avocado*" (P10) was interpreted as "chicken eggs" and "avocado", but could alternatively be chicken meat (e.g., "*chicken and rice 4 oz*" P10) and not a description of egg type. Other examples include "*salmon cakes . . .*" (P14), "*peanut butter muffin*" (P12), and "*banana tea*" (P9). Similarly, decomposed foods that had a food name followed by individual ingredients could be counted twice. For instance, "*pasta / 3 oz. edamame spaghetti, 4 variety tomato, .5 tbsp. olive oil, [. . .]*" (P6) was interpreted as general pasta as well as edamame spaghetti, tomato, and so forth.

Most descriptive inputs had food items where amount was specified, but amount interpretability depended on how it was described. Scale and numeric descriptions had 78.7% and 80.0% of inputs completely and correctly interpreted, while non-standard measures were correctly interpreted in about half of inputs (53.8%). Some of the non-standard measures could be occasionally understood and return estimated nutritional metrics, such as bowl, spoonful, bottle, plate and scoop. However, "handful" was not captured as a measure in any of its 23 occurrences. Inputs where the amount was not clarified tended to return a default scale measure estimated by serving size, such as "*cereal with milk*" (P13) being assumed as 1 cup each. Similarly, "serving" was also mapped to default measures, such as "*bacon .4 serving*" being record as 0.4 of unit "slice".

4.5.2 Classification of food images. Most images had at least one food composition identified by the Clarifai service (42/54) with probabilities above 0.8, versus CloudVision and Rekognition that correctly identified components in 22/60 and 21/60, respectively. However, the latter two services accurately identified background elements in 14 images (e.g., table, keyboard) and food containers in 36 images (e.g., plate, bowl), whereas Clarifai did not identify these elements at all. We based our analysis on food identification on Clarifai's results and background elements on results from CloudVision and Rekognition.

Participants frequently took photos of packages, wrapped foods, or uploaded stock photos of food items, representing a third of the images participants uploaded (19/60). Non-stock images of packages were mostly classified correctly (5/7), such as Figure 4d being classified as "*chocolate*" (1.00), "*candy*" (0.98), "*sweet*" (0.91). However, shape and color occasionally influenced the recognition, with the cookie snack in Figure 4l classified as "*beer*" (0.96), "*bacon*" (0.95), "*cake*" (0.92), or "*chips*" (0.83). 7/10 stock images had key components identified, such as Figure 4e that had "*oatmeal*" and "*cereal*" among high-confidence food items. However, the three other stock image inputs, two inputs of "*Dried mango slices*" and one of "*Apple Smoked Bacon*", had food names written on the package but failed to be correctly classified. This may be because pictures of the food items were not prominently displayed on the packaging, although other similar packages were classified correctly (e.g., Figure 4e). Neither of the 2 photos of wrapped foods were correctly identified, with food composition suggestions being influenced by the package and components of the background. For example, Clarifai incorrectly classified the wrapped sandwich in Figure 4a as "*chocolate*" or "*cake*" with high confidence (0.92 and 0.88), whereas CloudVision and Rekognition identified the "*aluminum foil*" (0.98) rather than the sandwich.

As expected, images with unclear food composition (6) were not well-classified by models. However, images of foods that had clear identifiable composition and that were not stock or packaged foods were mostly classified correctly (29/33). For example, Figure 4i was classified as "*salsa*" (0.97), "*corn*" (0.96), "*vegetable*" (0.95), "*tortilla chips*" (0.87), "*pepper*" (0.86), "*tomato*" (0.84), and "*chili*" (0.81). In contrast, Figure 4n also had tortilla chips and salsa but was incorrectly classified as "*bread*" (0.84) "*peanut*" (0.82), or "*peanut butter*" (0.81). This demonstrates that although generally images were correctly classified, inconsistencies around recognition make the method appear unpredictable and unexplainable.

Several background elements in 14 images were correctly identified, such as computer keyboards (e.g., Figure 3l, 0.99 probability), screens (e.g., Figure 4f, 0.92 probability), tables (e.g., Figure 4f, 0.64 probability), and even a kitchen oven in Figure 4m (0.57). Food

containers were also mostly identified (36/38), such as bowls (e.g., Figure 4g, 0.95 probability), plates (e.g., Figure 4b and 4i, 0.62 and 0.60 probabilities), cups (e.g., Figure 4j, 0.98 probability), and the blender in Figure 4m (0.98). However, the bowl in Figure 4f and the cake pan in Figure 4o were not detected, either because other background elements were identified with higher confidence (display 0.92, screen 0.92, monitor 0.92 . . .) or the food itself was recognized (chocolate 0.97, fudge 0.80, . . .).

5 DISCUSSION

Through deploying ModEat, we observed that when provided with a range of input modalities less constrained by recognition methods than current commercial tools and research prototypes, participants had high variance in how they preferred to describe their foods. Participants varied in how they described their foods collectively and individually, ranging from single to aggregated inputs with varying levels of specificity and detail, often describing food composition and varieties but sometimes created less specific entries. Participants also varied in how and whether they described amounts, either with numbers or formal scales but also with subjective references that made personal sense. Our study also suggests that automatic NLP for food description can identify food elements people describe fairly well, and classification services for image recognition can identify some food elements in most images. However, our findings also suggest that these automatic recognition services are optimized for identifying well-formed descriptions over free text, and plated food rather than the packages or stock photos participants often showed when journaling.

Reflecting on our results, we consider implications for designing both to mitigate and leverage the high levels of variance in how people prefer to describe foods. We also reflect on implications of automatic food interpretation.

5.1 Designing to Mitigate Variance in Food Descriptions

Prior work in personal informatics have suggested that technology can better support people's goals through customization or flexibility about what is recorded, allowing people to align self-tracking to their needs [7, 54]. However, we have also observed that supporting flexibility during data collection led to great variance in food description styles, some of which are not precise or introduce ambiguity. Ambiguity, in turn, can lead to challenges when data might want to be reviewed or reflected on later. This can be a particular issue for those with goals related to quantifying nutritional aspects of food consumption. Ambiguous food descriptions can lead to inaccuracy in food metrics, uncertainty about nutritional information, and ultimately not allowing for more precise reflection on progress toward a quantitative goal (e.g., Did I surpass my calorie budget? Have I consumed my protein quota for the day?). Likewise, completely unstructured inputs run the risk of introducing enough uncertainty that days or weeks later, people with awareness goals (e.g., learning about eating habits) might not be able to interpret their logs.

To mitigate ambiguity and uncertainty in logs, food journaling systems could encourage inclusion of food granularity, specificity, and amount by surfacing what was recognized and enabling correction or incrementation. Although recognition libraries often attempt to estimate nutrient values for ambiguous foods and those where amounts are unspecified, these values could be off from the reality of what a person consumed. Compared to gold-standard clinician-assisted 24- and 48-hour recalls, commercial food journals typically underestimate foods consumed [21, 22]. Current food databases show portion, calorie, and nutrient estimates based on what they search for, allowing them to edit or confirm prior to entry. Implementations of image recognition libraries or voice journaling could operate similarly, asking a person to confirm whether a food was correctly identified and how much was eaten.

For people with quantitative food goals, conversational journaling could enable prompting for greater specificity or clearer amounts to produce journals which more accurately represent what a person ate. We observed that generic food descriptions were often foods which could vary widely in calorie and nutrient information, such as "*pizza*" varying by toppings, slices eaten, and slice size. When someone with a goal that requires accurate metrics creates an entry that contains ambiguous characteristics, conversational journals could detect and interact to highlight the issue (e.g., report that an amount is missing) and offer suggestions to clarify or increase details of the entry (e.g., ask if the estimated or standard serving quantity for that food is accurate). This type of feedback could trigger more mindful consideration of foods consumed [10, 37], but requires careful consideration for balancing improving journal entry detail with burden [23] and feelings of judgment [31]. Conversational journaling could further adapt to people's use of non-standard amount measures. For instance, if someone uses a personally-meaningful reference point, the journal could work with the person journaling to jointly estimate portion size, such as comparing with an object with relatively standard proportions (e.g., a tennis ball) and remembering that estimate for future logs [20].

5.2 Designing to Allow for and Leverage Variance in Food Descriptions

Although uncertainty and inaccuracy is often seen as a negative in journaling and self-tracking [24], methods aimed to address variance could introduce burdens to the already-demanding food journaling domain. Allowing for flexibility or even treating variance in how people prefer to describe their food as a design opportunity could allow systems to support people's journaling goals without introducing unnecessary demands. Supporting variance may be particularly beneficial for people with awareness goals, who may have less of a need for a detailed or accurate record of what they ate.

While designing to effectively support the creation of accurate calorie or nutrient logs will ensure completeness and decrease variance, it has the downside of imposing structure on the data a person must enter. Beyond using conversational approaches or clarification questions to add detail about the food a person is eating, journals could aim to flexibly support adding different kinds of further detail. For example, a journal could provide open-ended fields for a person to include contextual information (e.g., whether eating alone or with others), information which could promote later

reminiscence (e.g., how they are feeling, a memory associated with that food), or anything else they wish the journal to know about their food. This could potentially balance desires for flexibility and detailed entries, leaving people to journal however they prefer and perhaps support awareness alongside quantitative journaling goals.

Past work has suggested that interpreting ambiguous journal entries can also provide an opportunity for deeper reflection around the circumstances under which the data was collected [3], potentially highlighting the social and cultural celebratory nature of food [47]. For instance, we observed that the presence of others during food journaling impacted how foods were described (e.g., aggregating foods in a single description). From this perspective, food description ambiguity can positively serve as a means of highlighting people's positive interactions with foods and encourage reflecting on around the circumstances which surrounded such an entry.

5.3 Implications of Automatic Food Interpretation

We found that commercial NLP and image classification services were reasonably successful in interpreting and identifying the foods that participants logged using their preferred strategies. Our results indicate that input structure influences NLP performance, with inputs with more specificity and standard amount descriptions more likely to be correctly recognized. We were surprised that most food descriptions were correctly interpreted and returned nutritional information for identified foods. While this might be sufficient for quantitative-focused goals, commercial NLP systems were unable to interpret the contextual cues participants occasionally put in in food descriptions describing their location or social circumstances. Likewise, non-standard food descriptions that reference routines (e.g., "same as lunch. . .") or subjective amount descriptions, were also a challenge for automatic inference of logs, but are valuable information that people wanted to record. Similarly, image recognition libraries traded off accuracy for identifying the foods in an image with accuracy for identifying contextual information, such as what room a person might be eating in or whether they are eating from a plate or another container.

Our results, as well as others, suggest that people intend to collect contextual information for later reflection [30], perhaps pointing towards an opportunity for recognition models that comprehend not only food items, but other data. Incorporating models specifically trained for recognizing food in text and images together with other classification models (e.g., optical character recognition, more general object recognition models, barcode recognizers, amount classifiers) could enable adding such context as well as supporting recognition. Even if not identified with high detail or accuracy, these models could help point to fun or personally meaningful experiences [47], such as surfacing restaurant names or household objects visible in communal dining. Further mining of information embedded in photo metadata or passively recorded (e.g., location, time) could further provide context to complement reminiscence and reflection for both quantitative and awareness goal groups.

While automation is typically leveraged for food tracking towards calorie or nutrient goals, surfacing contextual elements from food photos and descriptions might also be beneficial for people with mindfulness and behavior awareness goals. As food journals become easier to collect passively through automated sensing [11, 70, 89] or with lower journaling burden [23, 30], there are increased opportunities for reflecting on long-term logs. Photos and text descriptions can be difficult to aggregate, but automation can promote longer-term reflection or reminiscence by mining abstract concepts from these logs. For instance, people could be provided with a cloud of words with the names of frequent foods or food categories they journaled, or display a color gradient representing how the color of these foods has varied over time.

Food journaling can also benefit from image classification for increasing detail of food consumption. People with calorie and nutrient goals could leverage this by confirming identified foods in the image, adding further specificity about the ingredient makeup, and possibly clarifying amounts. Identified and confirmed foods could then be automatically searched for nutritional data in a database. Amounts could also be suggested based on contextual information present in the image, such as text on packages or food inside containers (e.g., plate, cup). This semi-automated approach might potentially lower journaling time and effort [58], while still promoting engagement [23].

6 CONCLUSION

In deploying ModEat, a lightweight food journaling technology prototype, we have identified that participant's strategies for describing foods had high variance, ranging from granular to aggregated inputs, and different levels of specificity and ways of describing amounts. We also observed that food descriptions or images could also be ambiguous and often not clear as to actual consumed amounts. The strategies which people use to create food logs were typically interpretable by recognition libraries, but were less successful for aggregated or less specific food inputs. Our findings point to opportunities for conversational food journaling to help mitigate variance by supporting adding further detail, but also for technology to leverage journaling variance to promote reminiscence. Leveraging automatic food interpretation can additionally lower journaling burden or add context, supporting increased value from long-term food logs.

ACKNOWLEDGMENTS

We thank Kimberly Flores and Yuqi Huai for helping to develop ModEat. We also thank Elizabeth Ankrah, Isil Oyguï, and Zhaoyuan Su for feedback on deployment piloting, Jong Ho Lee and David V. Nguyen for helping with analysis of some of the commercial services, and our participants. This research was supported in part by the National Science Foundation under award IIS-1850389.

REFERENCES

[1] 7 Best Image Recognition APIs. Retrieved 10 February, 2021 from https://nordicapis.com/7-best-image-recognition-apis/

[2] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. (2015). You Tweet What You Eat: Studying Food Consumption Through Twitter. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2015)*, 3197–3206. http://doi.org/10.1145/2702123.2702153

[3] Deemah Alqahtani, Caroline Jay, and Markel Vigo. (2020). The Role of Uncertainty as a Facilitator to Reflection in Self-Tracking. *Proceedings of the Conference on Designing Interactive Systems* (DIS 2020), 1807–1818. http://doi.org/10.1145/3357236.3395448

[4] Amazon Rekognition. Retrieved 10 February, 2021 from https://aws.amazon.com/rekognition/

[5] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. (2005). Analysis of Chewing Sounds for dietary monitoring. *Proceedings of the International Conference on Ubiquitous Computing (Ubicomp 2005)*, 56–72. http://doi.org/10.1007/11551201_4

[6] Adrienne H. Andrew, Gaetano Borriello, and James Fogarty. (2013). Simplifying Mobile Phone Diaries: Design and Evaluation of a Food Index-Based Nutrition Diary. *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth 2013)*, 260–263. http://doi.org/bbkk

[7] Amid Ayobi, Paul Marshall, and Anna L. Cox. (2020). Trackly: A Customisable and Pictorial Self-Tracking App to Support Agency in Multiple Sclerosis Self-Care. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2020)*, 1–15. http://doi.org/10.1145/3313831.3376809

[8] Barcode Lookup. Retrieved 10 February, 2021 from https://www.barcodelookup.com/

[9] Elizabeth Barrett-Connor. (1991). Nutrition epidemiology: How do we know what they ate? *American Journal of Clinical Nutrition*, 182–189. http://doi.org/10.1093/ajcn/54.1.182s

[10] Eric P.S. Baumer, Sherri Jean Katz, Jill E. Freeman, Phil Adams, Amy L. Gonzales, John Pollak, Daniela Retelny, Jeff Niederdeppe, Christine M. Olson, and Geri K. Gay. (2012). Prescriptive Persuasion and Open-Ended Social Awareness: Expanding the Design Space of Mobile Health. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2012)*, 475–484. http://doi.org/10.1145/2145204.2145279

[11] Abdelkareem Bedri, Gregory Abowd, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, and Thad Starner. (2017). EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 1(3), 1–20. http://doi.org/10.1145/3130902

[12] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. (2020). FitByte: Automatic Diet Monitoring in Unconstrained Situations Using Multimodal Sensing on Eyeglasses. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2020)*, 1–12. http://doi.org/10.1145/3313831.3376869

[13] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. (2015). Menu-match: Restaurant-Specific Food Logging From Images. *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV 2015)*, 844–851. http://doi.org/10.1109/WACV.2015.117

[14] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. (2019). Communication Breakdowns Between Families and Alexa. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2019)*, 1–13. http://doi.org/10.1145/3290605.3300473

[15] Johnna Blair, Yuhan Luo, Ning F. Ma, Sooyeon Lee, and Eun Kyoung Choe. (2018). OneNote Meal: A Photo-Based Diary Study for Reflective Meal Tracking. *AMIA Annual Symposium proceedings (AMIA 2018)*, 252–261. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371351/

[16] Marc Bolanos and Petia Radeva. (2016). Simultaneous Food Localization and Recognition. *Proceedings of International Conference on Pattern Recognition (ICPR 2016)*, 0, 3140–3145. http://doi.org/10.1109/ICPR.2016.7900117

[17] Virginia Braun and Victoria Clarke. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://doi.org/10.1191/1478088706qp063oa

[18] Lora E. Burke, Molly B. Conroy, Susan M. Sereika, Okan U. Elci, Mindi A. Styn, Sushama D. Acharya, Mary A. Sevick, Linda J. Ewing, and Karen Glanz. (2011). The Effect of Electronic Self-Monitoring on Weight Loss and Dietary Intake: A Randomized Behavioral Weight Loss Trial. *Obesity, 19*(2), 338–344. http://doi.org/10.1038/oby.2010.208

[19] Micael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome, and Matthieu Cord. (2018). Cross-Modal Retrieval in the Cooking Context: Learning Semantic Text-Image Embeddings. ning Semantic Text-Image Embeddings. *Proceedings of The Conference on Research & Development in Information Retrieval (SIGIR 2018)*, 35-44. http://doi.org/10.1145/3209978.3210036

[20] Beenish M. Chaudhry, Christopher Schaefbauer, Ben Jelen, Katie A. Siek, and Kay Connelly. (2016). Evaluation of a Food Portion Size Estimation Interface for a Varying Literacy Population. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2016)*, 5645–5657. http://doi.org/10.1145/2858036.2858554

[21] Juliana Chen, William Berkman, Manal Bardouh, Ching Yan Kammy Ng, and Margaret Allman-Farinelli. (2019). The Use of a Food Logging App in the Naturalistic Setting Fails to Provide Accurate Measurements of Nutrients and Poses Usability Challenges. *Nutrition*, *57*, 208–216. http://doi.org/10.1016/j.nut.2018.05.003

[22] Juliana Chen, Janet E. Cade, and Margaret Allman-Farinelli. (2015). The Most Popular Smartphone Apps for Weight Loss: A Quality Assessment. *Journal of Medical Internet Research (JMIR 2015)*, *3*(4), e104. http://doi.org/10.2196/mhealth.4334

[23] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A. Epstein, Felicia Cordeiro, Matthew Kay, Gregory D. Abowd, Tanzeem Choudhury, James Fogarty, Bongshin Lee, Mark Matthews, and Julie A. Kientz. (2017). Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing*, *16*(1), 74–84. http://doi.org/10.1109/MPRV.2017.18

[24] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. (2014). Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2014), 1143–1152*. http://doi.org/10.1145/2556288.2557372

[25] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. (2016). Characterizing Dietary Choices, Nutrition, and Language in Food Deserts Via Social Media. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2016), 1157–1170*. http://doi.org/10.1145/2818048.2819956

[26] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. (2018). Detecting Eating Episodes by Tracking Jawbone Movements with a Non-Contact Wearable Sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 2(1), 1–21*. http://doi.org/10.1145/3191736

[27] [27]Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A. Munson. (2019). Identifying and Planning for Individualized Change. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, *3*(1), 1–27. http://doi.org/10.1145/3314394

[28] Chia Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A. Munson. (2017). When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017), 1674–1687. http://doi.org/10.1145/3025453.3025747

[29] Clarifai's Food Model | AI Prediction of Specific Food in Meals. Retrieved 10 February, 2021 from https://www.clarifai.com/models/food

[30] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. (2015). Rethinking the Mobile Food Journal: Exploring Opportunities for Lightweight Photo-Based Capture. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2015), 3207–3216*. http://doi.org/10.1145/2702123.2702154

[31] Felicia Cordeiro, Daniel A. Epstein, Edison Thomaz, Elizabeth Bales, Arvind K. Jagannathan, Gregory D. Abowd, and James Fogarty. (2015). Barriers and Negative Nudges: Exploring Challenges in Food Journaling. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2015), 1159–1162*. http://doi.org/10.1145/2702123.2702155

[32] Alaina Darby, Matthew W. Strum, Erin Holmes, and Justin Gatwood. (2016). A Review of Nutritional Tracking Mobile Applications for Diabetes Patient Use. *Diabetes Technology & Therapeutics*, *18*(3), 200–212. http://doi.org/10.1089/dia.2015.0299

[33] Tamara Denning, Adrienne Andrew, Rohit Chaudhri, Carl Hartung, Jonathan Lester, Gaetano Borriello, and Glen Duncan. (2009). BALANCE: Towards a Usable Pervasive Wellness Application With Accurate Activity Inference. *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile'09)*, 5. http://doi.org/10.1145/1514411.1514416

[34] Pooja M Desai, Elliot G Mitchell, Maria L Hwang, Matthew E Levine, David J Albers, and Lena Mamykina. (2019). Personal Health Oracle: Explorations of Personalized Predictions in Diabetes Self-Management. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2019), 1–13*. http://doi.org/10.1145/3290605.3300600

[35] E-health application categories used by U.S. adults 2017 | Statista. Retrieved 10 February, 2021 from https://www.statista.com/statistics/378850/top-mobile-health-application-categories-used-by-us-consumers/

[36] Edamam Nutrition Analysis API. Retrieved 10 February, 2021 https://developer.edamam.com/edamam-nutrition-api

[37] Daniel A. Epstein, Felicia Cordeiro, James Fogarty, Gary Hsieh, and Sean A. Munson. (2016). Crumbs: Lightweight Daily Food Challenges to Promote Engagement and Mindfulness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2016), 5632–5644*. http://doi.org/10.1145/2858036.2858044

[38] Amazon says 100 million Alexa devices have been sold - The Verge. Retrieved 10 February, 2021 https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp

[39] Giannina Ferrara, Jenna Kim, Shuhao Lin, Jenna Hua, and Edmund Seto. (2019). A Focused Review of Smartphone Diet-Tracking Apps: Usability, Functionality, Coherence With Behavior Change Theory, and Comparative Validity of Nutrient Intake and Energy Estimates. *JMIR mHealth and uHealth*, *7*(5), e9232. http://doi.org/10.2196/mhealth.9232

[40] FoodData Central - USDA. Retrieved 10 February, 2021 from https://fdc.nal.usda.gov/api-guide.html

[41] Google's Cloud Vision API. Retrieved 10 February, 2021 from https://cloud.google.com/vision/docs/

[42] Google's "smart" Food Diary is Actually Kind of Dumb. Retrieved 10 February, 2021 from https://www.theverge.com/2015/6/2/8707851/google-calories-food-photos-im2calories

[43] Nanna Gorm and Irina Shklovski. (2017). Participant Driven Photo Elicitation for Understanding Activity Tracking: Benefits and Limitations. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2017)*, 1350–1361. http://doi.org/10.1145/2998181.2998214

[44] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. (2018). Activity Tracking in Vivo. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2018)*, 1–13. http://doi.org/10.1145/3173574.3173936

[45] Alexandros Graikos, Vasileios Charisis, Dimitrios Iakovakis, Stelios Hadjidimitriou, and Leontios Hadjileontiadis. (2020). Single Image-Based Food Volume Estimation Using Monocular Depth-Prediction Networks. *Universal Access in Human-Computer Interaction. Applications and Practice (HCII 2020)*, 532–543. http://doi.org/10.1007/978-3-030-49108-6_38

[46] Andrea Grimes, Martin Bednar, Jay David Bolter, and Rebecca E Grinter. (2008). EatWell: Sharing Nutrition-Related Memories in a Low-Income Community. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2008)*, 87-96. http://doi.org/10.1145/1460563.1460579

[47] Andrea Grimes and Richard Harper. (2008). Celebratory Technology: New Directions for Food Research in HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, 467–476. http://doi.org/10.1145/1357054.1357130

[48] William D. Heizer, Susannah Southern, and Susan McGovern. (2009). The Role of Diet in Symptoms of Irritable Bowel Syndrome in Adults: A Narrative Review. *Journal of the American Dietetic Association*, *109*(7), 1204–1214. http://doi.org/10.1016/j.jada.2009.04.012

[49] Uta Hinrichs and Sheelagh Carpendale. (2011). Gestures in The Wild: Studying Multi-Touch Gesture Sequences on Interactive Tabletop Exhibits. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*, 3023–3032. http://doi.org/10.1145/1978942.1979391

[50] Jack F. Hollis, Christina M. Gullion, Victor J. Stevens, Phillip J. Brantley, Lawrence J. Appel, Jamy D. Ard, Catherine M. Champagne, Arlene Dalcin, Thomas P. Erlinger, Kristine Funk, Daniel Laferriere, Pao Hwa Lin, Catherine M. Loria, Carmen Samuel-Hodge, William M. Vollmer, and Laura P. Svetkey. (2008). Weight Loss During the Intensive Intervention Phase of the Weight-Loss Maintenance Trial. *American Journal of Preventive Medicine*, *35*(2), 118–126. http://doi.org/10.1016/j.amepre.2008.04.013

[51] Eunkyung Jo, Hyeonseok Bang, Myeonghan Ryu, Eun Jee Sung, Sungmook Leem, and Hwajung Hong. (2020). MAMAS: Supporting Parent - Child Mealtime Interactions Using Automated Tracking and Speech Recognition. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1). http://doi.org/10.1145/3392876

[52] Jisu Jung, Kalina Yacef, Margaret Allman-farinelli, Judy Kay, Lyndal Wellard-Cole, Colin Cai, Irena Koprinska, and Margaret Allman-Farinelli. (2020). Foundations for Systematic Evaluation and Benchmarking of a Mobile Food Logger in a Large-scale Nutrition Study. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 4(2), 47.* http://doi.org/10.1145/3397327

[53] Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. (2017). TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, 2017-May, 6850–6863. http://doi.org/10.1145/3025453.3025480

[54] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. (2017). OmniTrack: A Flexible Self-Tracking Approach Leveraging Semi-Automated Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 1(3), 1– 28.* http://doi.org/10.1145/3130930

[55] Mandy Korpusik, Zachary Collins, and James Glass. (2017). Semantic Mapping of Natural Language Input to Database Entries Via Convolutional Neural Networks. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, 5685–5689. http://doi.org/10.1109/ICASSP.2017.7953245

[56] Mandy Korpusik and James Glass. (2017). Spoken Language Understanding for a Nutrition Dialogue System. *IEEE/ACM Transactions on Audio Speech and Language Processing*, *25*(7), 1450–1461. http://doi.org/10.1109/TASLP.2017.2694699

[57] Ian Li, Anind Dey, and Jodi Forlizzi. (2010). A Stage-Based Model of Personal Informatics Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2010), 1, 557–566.* http://doi.org/10.1145/1753326.1753409

[58] Brian Y. Lim, Xinni Chng, and Shengdong Zhao. (2017). Trade-off between Automation and Accuracy in Mobile Photo Recognition Food Logging. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2017)*, 53–59. http://doi.org/10.1145/3080631.3080640

[59] Lose It! - Weight Loss That Fits. Retrieved February 10, 2021 from https://www.loseit.com/

[60] Kai Lukoff, Taoxi Li, Yuan Zhuang, and Brian Y. Lim. (2018). TableChat: Mobile Food Journaling to Facilitate Family Support for Healthy Eating. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–28. http://doi.org/10.1145/3274383

[61] Lena Mamykina, Matthew E. Levine, Patricia G. Davidson, Arlene M Smaldone, Noemie Elhadad, and David J Albers. (2016). Data-driven health management: reasoning about personally generated data in diabetes with information technologies. *Journal of the American Medical Informatics Association*, *23*(3), 526–531. http://doi.org/10.1093/jamia/ocv187

[62] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. (2008). MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, 477. http://doi.org/10.1145/1357054.1357131

[63] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. (2019). A Survey on Food Computing. *ACM Computing Surveys*, *52*(5), 1–36. http://doi.org/10.1145/3329168

[64] Jimmy Moore, Pascal Goffin, Miriah Meyer, Philip Lundrigan, Neal Patwari, Katherine Sward, and Jason Wiese. (2018). Managing In-home Environments through Sensing, Annotating, and Visualizing Air Quality Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 2(3), 1–28. http://doi.org/10.1145/3264938

[65] Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin Murphy. (2015). Im2Calories: Towards an automated mobile vision food diary. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1233–1241. http://doi.org/10.1109/ICCV.2015.146

[66] MyFitnessPal: Calorie Counter, Diet & Exercise Journal. Retrieved February 10, 2021 from https://www.myfitnesspal.com/

[67] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z Gajos. (2011). PlateMate: Crowdsourcing Nutrition Analysis from Food Photographs. *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST 2011)*, 1-12. http://doi.org/10.1145/2047196.2047198

[68] NPR Study Says 118 Million Smart Speakers Owned by U.S. Adults - Voicebot.ai. Retrieved February 10, 2021 from https://voicebot.ai/2019/01/07/npr-study-says-118-million-smart-speakers-owned-by-u-s-adults/

[69] Nutritionix Nutrition API. Retrieved February 10, 2021 from https://www.nutritionix.com/business/api

[70] Hyungik Oh, Jonathan Nguyen, Soundarya Soundararajan, and Ramesh Jain. (2018). Multimodal Food Journaling. *Proceedings of the International Workshop on Multimedia for Personal Health and Health Care (HealthMedia 2018)*, 39–47. http://doi.org/10.1145/3264996.3265000

[71] Tauhidur Rahman, Alexander T. Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. (2014). BodyBeat: Amobile system for sensing non-speech body sounds. *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys 2014)*, 2–13. http://doi.org/10.1145/2594368.2594386

[72] RapidAPI - Top Nutrition APIs. Retrieved February 10, 2021 from https://rapidapi.com/collection/nutrition

[73] Jaime Ruiz, Yang Li, and Edward Lank. (2011). User-Defined Motion Gestures for Mobile Interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011), 197–206.* http://doi.org/10.1145/1978942.1978971

[74] Amaia Salvador, Michal Drozdzal, Xavier Giro-I-Nieto, and Adriana Romero. (2019). Inverse Cooking: Recipe Generation from Food Images. *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, 10445–10454. http://doi.org/10.1109/CVPR.2019.01070

[75] Jessica Schroeder, Jane Hoffswell, Chia Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. (2017). Supporting Patient-Provider Collaboration to Identify Individual Triggers Using Food and Symptom Journals. *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2017)*, 1726–1739. http://doi.org/10.1145/2998181.2998276

[76] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. (2018). "Hey Alexa, what's up?": Studies of In-Home Conversational Agent Usage. *Proceedings of the Conference on Designing Interactive Systems (DIS 2018)*, 857–868. http://doi.org/10.1145/3196709.3196772

[77] Katie A. Siek, Kay H. Connelly, Yvonne Rogers, Paul Rohwer, Desiree Lambert, and Janet L. Welch. (2006). When Do We Eat? An Evaluation of Food Items Input into an Electronic Food Monitoring Application. *2006 Pervasive Health Conference and Workshops*, 1–10. http://doi.org/10.1109/PCTHEALTH.2006.361684

[78] Spoonacular food API. Retrieved February 10, 2021 from https://spoonacular.com/food-api

[79] Zhida Sun, Sitong Wang, Wenjie Yang, Onur Yürüten, Chuhan Shi, and Xiaojuan Ma. (2020). "A Postcard from Your Food Journey in the Past": Promoting Self-Refection on Social Food Posting. *Proceedings of the Conference on Designing Interactive Systems (DIS 2020)*, 1819–1832. http://doi.org/10.1145/3357236.3395475

[80] Edison Thomaz, Aman Parnami, Irfan Essa, and Gregory D. Abowd. (2013). Feasibility of Identifying Eating Moments from First-Person Images Leveraging Human Computation. *Proceedings of the International SenseCam & Pervasive Imaging Conference (SenseCam 2013)*, 26–33. http://doi.org/10.1145/2526667.2526672

[81] Christopher C. Tsai, Gunny Lee, Fred Raab, Gregory J. Norman, Timothy Sohn, William G. Griswold, and Kevin Patrick. (2007). Usability and Feasibility of PmEB: A Mobile Phone Application for Monitoring Real Time Caloric Balance. *Mobile Networks and Applications*, *12*(2–3), 173–184. http://doi.org/10.1007/s11036-007-0014-4

[82] Yunan Wang, Jing Jing Chen, Chong Wah Ngo, Tat Seng Chua, Wanli Zuo, and Zhaoyan Ming. (2019). Mixed dish recognition through multi-label learning. *Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities (CEA 2019)*, 1–8. http://doi.org/10.1145/3326458.3326929

[83] Donald A Williamson, ; H Raymond Allen, ; Pamela, Davis Martin, Anthony J Alfonso, Bonnie Gerald, and Alice Hunt. (2003). Comparison of Digital Photography

to Weighed and Visual Estimation of Portion Sizes. *Journal of American Dietetic Associattion*, *103*, 1139–1145. https://doi.org/10.1016/S0002-8223(03)00974-X

[84] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. (2009). *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2009)*, 1083–1092. http://doi.org/10.1145/1518701.1518866

[85] Hui Wu, Michele Merler, Rosario Uceda-Sosa, and John R. Smith. (2016). Learning to Make Better Mistakes: Semantics-Aware Visual Food Recognition. *Proceedings of the International Conference on Multimedia (MM 2016)*, 172–176. http://doi.org/10.1145/2964284.2967205

[86] WW (Weight Watchers): Weight Loss & Wellness Help. Retrieved February 10, 2021 from https://www.weightwatchers.com

[87] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. (2010). Food Recognition Using Statistics of Pairwise Local Features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 2249–2256. http://doi.org/10.1109/CVPR.2010.5539907

[88] YouFood Photo Food Journal on the App Store. Retrieved February 10, 2021 from https://apps.apple.com/us/app/youfood-photo-food-journal/id719841416

[89] Shibo Zhang, Yuqi Zhao, Dzung Tri Nguyen, Runsheng Xu, Sougata Sen, Josiah Hester, and Nabil Alshurafa. (2020). NeckSense: A Multi-Sensor Necklace for Detecting Eating Activities in Free-Living Conditions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 4(2), 1–26*. http://doi.org/10.1145/3397313

[90] Jasmine K. Zia, Chia-Fang Chung, Jessica Schroeder, Sean A. Munson, Julie A. Kientz, James Fogarty, Elizabeth Bales, Jeanette M. Schenk, and Margaret M. Heitkemper. (2017). *The Feasibility, Usability, and Clinical Utility of Traditional Paper Food and Symptom Journals for Patients with Irritable Bowel Syndrome*. Neurogastroenterology & Motility, 29(2), e12935. http://doi.org/10.1111/nmo.12935