

From Reflection to Action: Combining Machine Learning with Expert Knowledge for Nutrition Goal Recommendations

Elliot G. Mitchell*
Department of Biomedical
Informatics, Columbia University

Elizabeth M. Heitkemper
School of Nursing, The University of
Texas at Austin

Marissa Burgermaster
Department of Population Health,
Dell Medical School, and Department
of Nutritional Sciences, The
University of Texas at Austin

Matthew E. Levine
Department of Computing and
Mathematical Sciences, California
Institute of Technology

Yishen Miao
Department of Molecular, Cellular,
and Developmental Biology,
University of California Santa Barbara

Maria L. Hwang
Science and Math, Fashion Institute of
Technology

Pooja M. Desai
Department of Biomedical
Informatics, Columbia University

Andrea Cassells
Clinical Directors Network (CDN)

Jonathan N. Tobin
Clinical Directors Network (CDN)
and The Rockefeller University

Esteban G. Tabak
Courant Institute of Mathematical
Sciences

David J. Albers
University of Colorado, Anschutz
Medical Campus, Section of
Informatics and Data Science,
Departments of Pediatrics,
Biomedical Engineering, and
Biostatistics and Informatics, and
Department of Biomedical
Informatics, Columbia University

Arlene M. Smaldone
School of Nursing, Columbia
University

Lena Mamykina
Department of Biomedical
Informatics, Columbia University

ABSTRACT

Self-tracking can help personalize self-management interventions for chronic conditions like type 2 diabetes (T2D), but reflecting on personal data requires motivation and literacy. Machine learning (ML) methods can identify patterns, but a key challenge is making actionable suggestions based on personal health data. We introduce GlucoGoalie, which combines ML with an expert system to translate ML output into personalized nutrition goal suggestions for individuals with T2D. In a controlled experiment, participants with T2D found that goal suggestions were understandable and actionable. A 4-week in-the-wild deployment study showed that receiving

goal suggestions augmented participants' self-discovery, choosing goals highlighted the multifaceted nature of personal preferences, and the experience of following goals demonstrated the importance of feedback and context. However, we identified tensions between abstract goals and concrete eating experiences and found static text too ambiguous for complex concepts. We discuss implications for ML-based interventions and the need for systems that offer more interactivity, feedback, and negotiation.

CCS CONCEPTS

• **Applied computing** → Life and medical sciences; Consumer health; • **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI; Ubiquitous and mobile computing.

KEYWORDS

Personal Informatics, Machine learning, Goal setting, Diabetes self-management

ACM Reference Format:

Elliot G. Mitchell, Elizabeth M. Heitkemper, Marissa Burgermaster, Matthew E. Levine, Yishen Miao, Maria L. Hwang, Pooja M. Desai, Andrea Cassells,

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '21, May 08–13, 2021, Yokohama, Japan

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8096-6/21/05...\$15.00

<https://doi.org/10.1145/3411764.3445555>

Jonathan N. Tobin, Esteban G. Tabak, David J. Albers, Arlene M. Smaldone, and Lena Mamykina. 2021. From Reflection to Action: Combining Machine Learning with Expert Knowledge for Nutrition Goal Recommendations. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 08–13, 2021, Yokohama, Japan*. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3411764.3445555>

1 INTRODUCTION

Self-managing chronic conditions like type 2 diabetes (T2D) presents continual burden because it impacts countless choices individuals make in their daily lives [11]. Making healthy choices requires literacy and sustained motivation [11], and self-management is further complicated by the need for reflection and self-discovery due to high individual differences: the same choices can have profoundly different health impacts for different individuals [2, 70, 103]. These challenges contribute to growing health disparities; low income and minority communities have higher prevalence and worse outcomes from chronic diseases and lower access to resources like diabetes education [15, 46, 80].

Mobile health (mHealth) technologies can support self-management by leveraging behavior change techniques to help scaffold and support daily choices [55, 60]. Mobile and wearable devices also enable the collection of personal health data—such as diet and physical activity—which can lead to insights about the relationship between an individual’s behavior and their health state [23, 35, 71]. The field of personal informatics examines the use of personal data for reflection and increased self-knowledge, which could lead to improved health [64].

Despite this potential, many data-driven health interventions suffer from high user burden and low adoption [21, 61, 64]. The majority of interventions that incorporate self-tracking focus on viewing, visualizing, or reflecting on personal data. These approaches place the burden on individuals to derive insights from their data and determine how to change their behavior [47, 66]. As a consequence, individuals with low technology and health literacy, who are most impacted by chronic diseases, are least equipped to reap the benefits [98, 99].

One approach to help individuals more easily derive insights from their data is to apply *machine learning* (ML) to find patterns and make predictions. Recent research initiatives have demonstrated high accuracy in broadly health-related tasks [44, 85]. ML methods can be applied to personal health data to find patterns of association between multiple streams of self-tracking data [6] or forecast changes in blood sugar levels [27].

However, incorporating ML into personal health applications has its own challenges. Interpreting the output of an ML algorithm also requires knowledge and skills, and can be just as challenging as exploring self-tracking data. What’s more, even if ML can identify insightful patterns, those patterns may not be sufficient to help an individual understand how to change their behaviors: they may not be *actionable* if there is no information about what an individual can do to change or mitigate the unwanted outcomes [6, 47]. For example, an identified correlation between weather and physical activity may be less actionable without specific suggestions for how to stay active on rainy days [6]. Similarly, a prediction of high blood sugar may be less actionable without explaining what contributed to the forecast or how to mitigate it [27]. Generating suggestions

that inform individual action is the heart of the field of *recommender systems* (RecSys) [87, 95]. However, even for health-aware RecSys, ML is used to infer preferences, rather than the health impact of different choices; the health constraints for recommendations are assumed, not learned with ML [32, 83]. Other recent work has sought to incorporate recommendations based on ML-derived insights from self-tracking data, but were limited to an individual’s own past meals and therefore lacked variability [103], or relied on user’s self-perceptions of what behaviors impact health and were therefore unsurprising and less useful to users [47]. Thus, there is a need for new approaches to translating inferences achieved with ML into recommendations that can guide individuals’ action.

To address these research gaps, we developed an approach to couple ML inferences with an expert system in order to generate actionable recommendations; we used this approach to design a system called GlucoGoalie that makes personalized suggestions for nutritional goals for individuals with T2D. GlucoGoalie uses ML to identify patterns in self-tracking data—meals and BG levels captured with the GlucoGoalie smartphone app—regarding the relationship between nutrition and change in BG after meals. Furthermore, GlucoGoalie relies on a rule-based expert system to translate ML output into actionable support by generating natural language recommendations for nutrition goals in order to improve BG levels. These goals reflect both individual patterns identified with ML and expert knowledge regarding ways to improve BG management. Goal setting is a common approach to behavior change interventions, with demonstrated efficacy [30, 74, 75]. Each personalized goal is a suggestion to increase or decrease the amount of a macronutrient in meals, or to replace one macronutrient with another. Finally, GlucoGoalie helps individuals work towards achieving their goals by asking them to self-assess prospective meals on their consistency with selected goals during logging and to review a summary of goal achievement.

We evaluated GlucoGoalie in two complementary studies conducted with individuals with T2D recruited from economically disadvantaged communities. First, in a controlled experiment, we assessed whether participants understood goal recommendations generated by GlucoGoalie, and whether they could act on them to compose meals that were consistent with the recommendations using paper food cutouts. Second, we sought to understand the experience of using ML-derived health goals in-the-wild with a 4-week deployment study.

The controlled experiment showed that participants understood goals and chose the correct meal for a given goal with 89% accuracy when meals were accompanied by nutritional labels (only 49% accuracy with meal images that did not include labels). When composing meals using paper food cutouts, participants chose meals that were consistent with the direction of their selected goal 67% of the time (68 of 102; $p < 0.001$). This suggested feasibility of using computationally generated goal recommendations to inform individuals’ meal choices.

In the deployment study, participants found GlucoGoalie easy to use, fun, and engaging. Participants’ self-discovery, which is typical for self-tracking apps, was further augmented and informed by the experience of receiving suggestions for personalized goals. How they chose which goal to follow highlighted the importance and multifaceted nature of personal preferences in nutrition

recommendations. For example, participants looked for goals that were consistent with both their food preferences and their desire for variety in their diets. When working towards achieving their selected goals, participants described a desire for additional feedback on their progress. Moreover, they alluded to multiple contextual factors, such as their level of physical activity, previous meals, or even the season, that rendered certain goals as more or less useful in their daily lives. In addition, participants wished for a greater connection between goals that focused on their meals and their broader goals in life and health, such as reducing the need for medication. At the same time, some participants described difficulty understanding or acting on goal recommendations. Static text was sometimes ambiguous in communicating the goal's meaning, and because goals were based on abstract concepts like macronutrients, some participants had difficulty connecting goals to their concrete meal decisions.

This work contributes to a newly emerging understanding of how ML-based methods can be used to make actionable suggestion to support self-management of chronic conditions. Specifically, it outlines a particular approach to translating ML inferences into actionable recommendation by coupling ML with a rule-based expert system. We discuss ways in which systems intended to support action can also support reflection, describe how our approach to personalizing based on health constraints is complementary to existing approaches to ML-based personalization, and argue for the need for systems that offer more interactivity, feedback, and negotiation.

2 RELATED WORK

2.1 Diabetes Self-Management and Personalization

In chronic conditions like type 2 diabetes (T2D), daily lifestyle behaviors, including food choices and physical activity, impact short- and long-term health status. A key goal in T2D self-management is keeping blood glucose (BG) levels within target ranges. Dietary decisions have a direct impact on BG after each meal, but there is a high degree of variability between and within individuals. The same meal can have wildly different glycemic impacts across individuals [103], and the American Diabetes Association (ADA) recommends 1-on-1 counseling with diabetes educators to determine personalized nutrition goals and macronutrient targets [2]. Notably, it can be difficult for individuals to anticipate the impact of a particular meal on BG, even for experts [66].

2.2 Personal Informatics

Technologies for health and wellness have been a vibrant research area within HCI for several decades [55] with research focusing on persuasive behavior change interventions [81], health-focused games [41], and online health communities [68], among many other areas.

In addition to these more general interventions, many previous investigations have specifically focused on approaches to leveraging data collected with self-tracking. *Personal informatics* refers to a class of interactive technologies that help users to collect data about their behaviors and health, explore those data for patterns and trends, and engage in self-discovery, which is important for conditions with high individual variability like T2D [18, 64]. Li and

colleagues described five stages of the self-discovery process with personal informatics: preparation, collection, integration, reflection, and action [64]. While the first few stages focus on self-tracking of personal data, the latter stages—reflection and action—are about how individuals' beliefs and behaviors change in light of those data.

When it comes to reflection and action, the lion's share of previous research has focused on supporting *reflection* as a means towards greater self-knowledge [5, 33, 57, 65, 67]. Typical personal informatics systems rely on users to actively explore their data, which may require considerable literacy and effort [37, 66]. This approach, while appropriate for highly engaged users with available time and appropriate skills and knowledge [18, 64, 65], may present considerable challenges to the larger population with chronic disease, including individuals with low literacy and numeracy [14, 59]. Attempts to reduce the burden of reflection have focused on lightweight logging approaches that may still spur reflection, like photo or audio journals [25, 33, 40]. Other work has endeavored to support reflection on personal data through collaboration with peers or experts [19, 53, 79, 82], or creates more scaffolding for self-discovery by structuring self-experimentation [26, 50, 51]. Yet others have applied ML to identify patterns in multiple streams of self-tracking data and present them to users [6], or to make predictions about future BG levels based on a meal [27], but these approaches still fall short of direct support for action [47]. Notably, *EmotiCal* makes predictions about an individual's future mood, accompanied by recommendations for activities that might improve mood, but those recommendations are based on user-perceptions of what will improve mood, not direct inference from self-tracking data, and were therefore unsurprising and less useful to users [47]. In our research, we aim to build upon these previous efforts and use ML not only for arriving at inferences but adapt ML with an expert system to provide direct support for action by generating nutritional goal recommendations.

2.3 Data-Driven Goal Setting For Behavior Change

Goal setting is a common element in behavior change interventions [72] with demonstrated efficacy in mHealth applications [30]. Research in HCI has identified a number of important characteristics for goal setting applications including the need for autonomy in allowing users' input in choosing their own health goals, as well as the importance of helping individuals track progress in goal attainment with a visual summary [24]; this prior work informed design decisions for GlucoGoalie.

Many goal setting interventions help users articulate their health goals or choose from a set of options. Some approaches have sought to automate aspects of the goal-setting process, for example through interactive dialogs [9], or with adaptive goals [58]; however, these approaches tend to focus on a single dimension of tracking, like step counts or calories, and may not be applicable for more complex nutritional choices [83]. Recent work by Schroeder [92] highlighted an interconnection between an individual's goals and the types of self-tracking they engage in. In line with goal-directed tracking, GlucoGoalie uses self-tracking of meals and BG to inform personalized nutrition goals, and to assess progress in achieving selected goals over time through tracking.

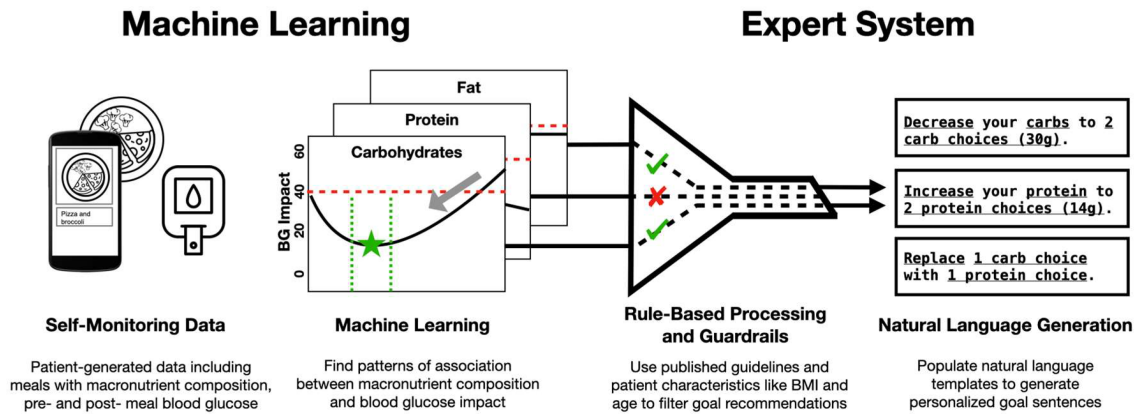


Figure 1: An overview of the pipeline for generating personalized goal recommendations in GlucoGoalie.

2.4 Recommender Systems

Another set of methods and interventions for personalized support based on individual user data are recommender systems (RecSys) [87]. RecSys are especially common in the consumer space, for example recommending items for purchase on an e-commerce site, or videos on a streaming platform. In the case of health, however, applying RecSys approaches is more challenging [90]. RecSys aim to personalize recommendations based on user preferences but can have unintended consequences because what individual prefers and what is healthy may be at odds. An emerging research area aims to incorporate health constraints into RecSys methods [31, 32, 89, 102] based on user-specified diet preferences [102], weight loss goals [31], or population-level guidelines [83, 89]. While these approaches make important strides towards healthy recommendations, it is important to note that even these systems use ML primarily to infer user preferences, and impose assumed *a priori*, rather than learned, constraints on health [88]. Little work has explored how to learn personalized diet constraints or goals from self-tracking data.

3 GLUCOGOALIE

We designed GlucoGoalie through a user-centered design process building on our prior research with individuals with T2D from a predominantly Black and Latino economically disadvantaged community [86]. Figure 1 presents an overview of the pipeline for generating personalized goals. GlucoGoalie’s goal-generating engine includes two main components: a machine learning algorithm for detecting patterns of association between nutrition in meals and changes in BG levels, and an expert system that uses expert knowledge to generate recommendations for nutritional goals in order to improve BG levels. We describe these in more detail below.

3.1 Approach to Goal Setting

Our aim was to generate personalized recommendations for nutritional goals that can be actionable and easily understood by individuals with mixed levels of literacy. One of the key decisions in the design of GlucoGoalie was regarding the level of specificity in

nutritional goals. We worked with a group of Certified Diabetes Educators (CDEs, $n=3$) to formulate goals that are consistent with the ones used in typical diabetes education and that focus on changes to macronutrient composition of meals. We made this choice for three main reasons. First, the macronutrient composition of a meal is directly related to its impact on BG, but the specifics of the relationship vary between individuals [36]. Second, nutrition education in diabetes emphasizes macronutrients to help individuals think flexibly about the nutritional composition of similar foods [101]. Third, using macronutrients as features has advantages for machine learning, offering a denser, low-dimensional feature representation compared to other representations like the specific food items in a meal. We worked with CDEs to create templates for goals that could be populated by an ML algorithm and identified three types of changes to meal composition that could impact post-meal BG: increase the amount of macronutrient, decrease the amount, or replace one macronutrient with another. Goals are meal-level because the balance of each meal has its own impact on BG, making day-level goals (e.g., daily calories) less appropriate. See Table 1 for a selection of goals.

3.2 Machine Learning

The high-level aim of the ML problem was to infer the relationship between an individual’s nutrition choices and changes in their BG levels after meals. The features in the ML problem are the meals a user has logged, specifically the grams of carbohydrates, protein, and fat. The outcome of interest—change in BG after a meal—is the difference between self-reported BG before the meal, compared with 2 hours after, which is the clinical standard [4]. The ML method to find patterns of association between nutrition and BG was based on Attributable Components Analysis (ACA), a non-parametric method for estimating the conditional expectation of a quantity of interest based on a set of covariates [96]. Because self-monitoring data are manually entered by users, there are often a small number of data points that are prone to include errors and outliers. These characteristics pose challenges for ML, and ACA has advantages over other methods like linear regression because it is able to capture non-linear relationships, is less sensitive to erroneous data points, and more effectively estimates uncertainty [73].

Table 1: A selection of nutritional goals available in GlucoGoalie. Generic goals are available for all users from when they first use the app. Personalized goals are recommended for an individual based on ML-based analysis of recorded meals and blood glucose readings. Underlined words are personalized for each user based on their data. Note: A food “choice” is a unit similar to a serving size that identifies servings of different foods with similar macronutrient compositions.

Type	Title	Description
Personalized	<u>Decrease your carbs to 2½ carb choices</u>	For <u>high carb lunches</u> , decrease your carbs to be about <u>2½ carb choices (38g)</u> . An example of 1 carb choice is 1 slice of whole wheat toast, $\frac{1}{3}$ cup of plantains, or $\frac{1}{3}$ cup of brown rice.
Personalized	<u>Increase your protein to 3 protein choices</u>	For <u>low protein dinners</u> , <u>increase your protein</u> to be about <u>3 lean protein choices (21g)</u> . An example of 1 lean protein choice is 1 ounce of lean ground beef, $\frac{1}{2}$ cup of tofu, or 1 ounce of chicken breast.
Personalized	Replace <u>2 carb choices</u> with <u>2 protein choices</u>	For <u>high carb dinners</u> , replace <u>2 carb choices</u> with <u>2 lean protein choices</u> . For example, replace $\frac{2}{3}$ cup of rice with 2 ounces of ground turkey or 2 ounces of tilapia.
Generic	Choose whole fruits	Choose whole fruits instead of fruit juices. For example, have a whole orange, an apple, or a cup of berries with your meals.
Generic	Choose plant proteins	Include proteins that come from plants, such as beans, nuts and seeds, and legumes. For example, choose a cup of beans, a handful of peanuts, or a cup of lentils to add protein to your meal.

3.3 Expert System Interpretation and Guardrails

While ML can identify patterns in the relationship between meals and BG, these patterns alone are not sufficient to inform behavior. In a series of 10 sessions, we worked with CDEs to establish rules for interpreting the ML output and translating it into goal recommendations. For example, GlucoGoalie suggests goals only if ML infers patterns with an expected increase in BG above a clinically significant threshold (40 mg/dl). In addition, CDEs pointed out that some automatically generated recommendations might be inappropriate irrespective of their impact on BG, for example a goal to eat 100g of fat in a single meal. To mitigate this concern, we added a set of guardrails to filter out extreme recommendations based on population-level nutrition guidelines.

In co-designing the goal templates with CDEs, we also sought to formulate goals such that they could be understood and acted upon by individuals, even those with low nutrition literacy. Because we could not assume nutrition knowledge, we embedded necessary information within the goals themselves. First, each goal includes three examples of concrete foods rich in a target macronutrient. Examples are drawn from a knowledge base created using an ADA resource [101]. To increase their relevance, examples were selected from meal logs captured by participants of a prior self-tracking study; these participants were recruited from a similar population and captured their regular meals for 2-5 weeks, thus creating a rich collection of meals. Second, we considered multiple approaches to describing target macronutrient amounts, including standard units like grams, heuristics like fists and thumbs, or even the proportions of a plate covered with different types of foods, an approach consistent with ChooseMyPlate [97]. However, we dismissed visualizing proportions on a plate due to their lack of precision (15g of rice could be gathered together in a ball or spread thinly over the entire plate). Instead, we opted for the ADA-endorsed language of food “choices,” a system meant to simplify nutrition education [101]. A

food “choice” is a unit similar to a serving size; it identifies servings of different foods with similar macronutrient compositions. For example, 1 carbohydrate choice is 15 grams, which could be 1 slice of toast or $\frac{1}{3}$ cup of rice. In addition, because “choices” are based on grams, the standard unit on food labels, each goal also includes the target amount in grams.

3.4 The GlucoGoalie App

To explore individuals’ perceptions and experience receiving personalized goal suggestions in-the-wild, we included them in a custom smartphone application with logging and goal-setting functionality.

GlucoGoalie helps individuals set goals for improving their diet and work towards achieving these goals. Users begin by choosing one or more nutritional goals from a list in the app (see Table 1 for a selection of goals). To promote engagement with the application before users have tracked enough meals to receive personalized goal recommendations, all users choose from the same set of “generic” goals at the outset. Each generic goal describes a generally healthy behavior, and was developed by experts in nutrition and diabetes [22]. Twice per week, GlucoGoalie analyzes the data of each user with at least 8 meals to generate personalized nutrition goals, described above. If new goals are available, GlucoGoalie sends a push notification, and users can view the new, personalized recommendations and choose any they wish to follow (Figure 2d).

Within the app, users can log their meals and enter their pre-meal BG. Two hours after the meal, GlucoGoalie sends a push notification reminder to enter a post-meal reading. To simplify the logging process, users log meals by taking a picture of the meal and typing a free text description (Figure 2a). Macronutrient data are entered by a team of Registered Dietitians (RDs) who assessed each meal following a standard protocol based on the USDA nutrition database [52], but similar results could be attained via crowdsourcing [78]. To keep goals as a central part of the experience and promote

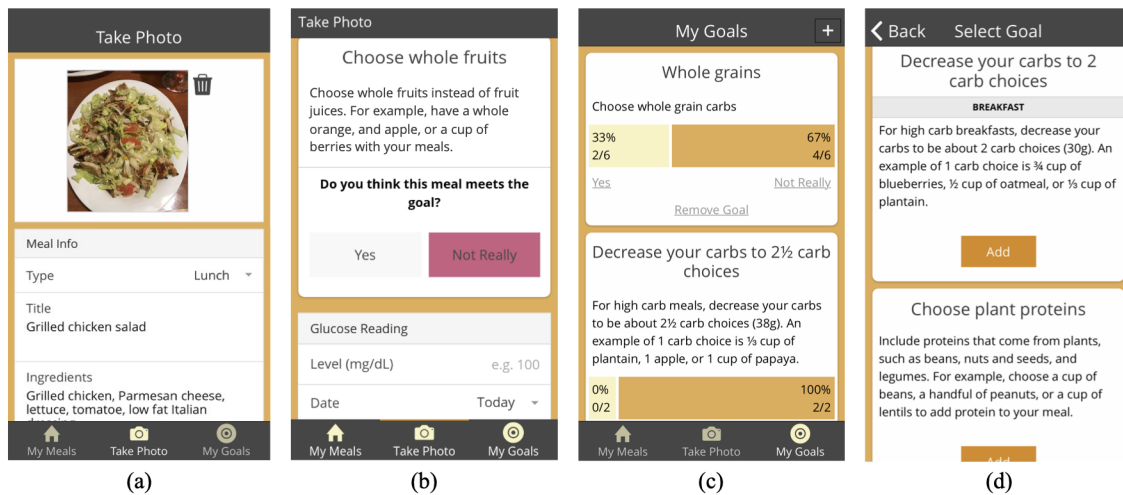


Figure 2: The GlucoGoalie mobile application. (a) Logging a meal with a photo and free text description. (b) Users self-assess whether they met their chosen goals. (c) A summary of goal achievement. (d) Reviewing and choosing new personalized goals to work on after receiving a push notification.

accountability, GlucoGoalie prompts users to assess whether their meal fits with each of their chosen goals while logging with either “Yes” or “Not Really” (Figure 2b); “Not Really” was identified as a preferred and less judgmental option than “No” during user-centered design. Users can view their current goals, remove or choose new goals, and review a summary of goal attainment in the My Goals section of the app (Figure 2c).

4 STUDY 1: CONTROLLED EVALUATION

In the first evaluation study, we sought to assess whether the types of personalized goals generated by the system would be *understandable* and *actionable* for individuals with T2D.

4.1 Controlled Evaluation Methods

4.1.1 Participants and Procedure. Participants were recruited from two types of health centers in a major US city: 1) a Federally Qualified Health Center (FQHC) in Brownsville Brooklyn, and 2) clinics affiliated with Columbia University Irving Medical Center. To be included, participants needed to be between 18- and 65-years-old with a self-reported diagnosis of type 2 diabetes (T2D) and proficient in English. After collecting consent, participants received a 10-minute, in-person nutrition training introducing the concept of food “choices” and reviewing macronutrients. Participants then completed the following three tasks:

Task 1: Goal Comprehension. To assess whether participants understood the content of the goals output by GlucoGoalie, participants were presented with an example goal “for a friend with diabetes,” and asked to choose which of two meals were a better fit with the goal. Meals were presented as a free text description with a nutrition label in the style of Facts Up Front [105] (see Supplementary Figure A); nutritional labels were included to ensure that this task was testing *comprehension of goals*, rather than individuals’ ability to assess nutritional composition of meals. This task was repeated twice.

Task 2: Goal/Image Matching. Because many meals are cooked at home, we included a second task to test comprehension of goal sentences using example meals *without* nutritional labels. Participants were again asked to choose which of two meals was a better fit with a presented, example goal, but could see only the meal image and description, with no macronutrient information (see Supplementary Figure B). Meal images were selected from a dataset collected in ongoing research with individuals from a similar population. Meal pairs were chosen to include similar ingredients but vary in macronutrient content; the incorrect answer was at least 1 macronutrient “choice” different from the correct answer, and the difficulty varied across scenarios. To account for higher variability in meal images, this task was repeated for eight goal/meal-pair combinations.

Task 3. Meal Choice: The Virtual Buffet. To assess whether goals were actionable and participants could follow them, we simulated the process of choosing meals with a “virtual buffet” made up of food image cutouts (see Supplementary Figure C). Participants then received a tailored goal and were asked to choose additional meals with that goal in mind. Working

1-on-1, researchers asked participants to use the food cutouts to assemble a baseline meal for each type of meal (breakfast, lunch, and dinner) that was “closest to what you would normally eat.” Importantly, there were multiple copies of each food item so participants could vary the amount of each food. Images were labeled with an amount in cups, tablespoons, or ounces, but never choices or grams (the units included in the goals themselves). We used images from a web-based resource [76] and our ongoing research and sought to include common staples like bacon and eggs as well as culturally relevant foods like fried plantains.

The baseline meals were used to identify goals that would require participants to vary from their typical macronutrient behavior by 1 to 1.5 macronutrient choices. For example, if a participant’s baseline

Table 2: Participant demographics for both studies

Demographics	Study 1: Controlled Experiment	Study 2: Deployment Study
N Enrolled (Incl. in Analysis)	19 (15)	10 (8)
Sex	80% Female	71% Female
Ethnicity	47% Hispanic	86% Hispanic
Race	17% White 42% Black 41% Other/Refused	43% White 29% Black 29% Other/Refused
Age	54 ± 9 years	55.7 ± 9.5 years
Body Mass Index (BMI)	37.4 ± 13.9	41.8 ± 14.4
Median Household Income	\$20,000-\$39,999	\$40,000-\$59,999
Median Education Level	Some College	Some College

meal had 3.5 carb choices and 1 protein choice, they would receive two goals: one to decrease carbs to 2 choices and another to increase protein to 2 choices. Participants chose one of the two goals, and then assembled “breakfast,” “lunch,” and “dinner” for three days in a row, with the chosen goal in mind (9 total meals). Researchers tallied the chosen food items to calculate nutrient compositions. During the 1-on-1 activity, researchers made note of participant comments and feedback, for example, questions about missing or inappropriate food items.

4.1.2 Data Analysis. For the goal comprehension and goal/image matching tasks, we calculated binary accuracy as a percentage ($\#correct / [\#correct + \#incorrect]$). For the “virtual buffet” experiment, we analyzed the data in two dimensions: direction and accuracy. First, we examined whether participants’ meals were consistent with the *direction* of their chosen goal. For example, if the goal was to increase protein to 2 choices at lunch, we assessed whether subsequent lunches had more protein than baseline. A binomial test was used to determine whether performance was better than chance. Second, accuracy in meeting the goal target was measured with mean absolute error between participant choices and the goal target. For example, if the target was 2 choices, we assessed how close participant’s meals were to the target, on average.

To synthesize participants impressions from their comments during the study, research met to debrief and aggregate notes in a series of meetings to summarize key themes.

4.2 Controlled Evaluation Results

4.2.1 Participants. We recruited and enrolled a total of 19 participants, including 10 from a Federally Qualified Health Center, and 9 from university-affiliated clinics. Four participants were excluded because of a data collection error for a total of 15 participants included in the analysis. As seen in Table 2, participants were predominantly female, and Black or Hispanic. Most were overweight or obese (body mass index ≥ 25).

4.2.2 Results. For the goal comprehension task, when choosing which of the two nutrition labels met a given goal participants were correct 89% (SD = 21%) of the time. When choosing which of two meal images was a better match with a goal, participants chose the correct meal 49% (SD = 25%) of the time. When composing meals at the virtual buffet, meals were consistent with the direction of

chosen goals 67% (68 of 102) of the time, significantly more than chance per a binomial test ($p < 0.001$). There was no difference in the percentage of meals consistent with the direction of chosen goals by meal type, macronutrient, or direction of goal. At the same time, there was a high degree of variability in precisely meeting the goal target. Meals were an average of 0.83 (SD = 0.56) “choices” away from the goal target. For example, given the goal “reduce carbs to 2 choices (30g),” participants were an average of $\frac{4}{5}$ carb choices (12g), from the target.

We identified two key themes in the comments made by participants during and after the virtual buffet activity. First, most participants commented on the limited selection of food items to choose from. Many recounted what they would normally eat, which was sometimes missing, for example oatmeal at breakfast. Usually, participants were able to select items they do eat from the available choices. Second, when choosing which goal to follow, participants often stated that they understood the goals. However, use of “choices” as a unit led to confusion, and some participants expressed uncertainty about how much food to take. Participants interpreted “2 choices” to mean two different food items, regardless of the amount (e.g., rice and bread) as opposed to a measurable quantity (e.g., $\frac{2}{3}$ cup of rice), as intended.

5 STUDY 2: DEPLOYMENT STUDY

In the second study, we sought to better understand participants’ experience receiving and following personalized goal recommendations from an app, based on their own self-tracking data, and conducted a 4-week deployment study with GlucoGoalie. Due to the timeline of this externally funded project, we did not update the design between the two studies.

5.1 Deployment Study Methods

5.1.1 Participants and Procedure. Participants were recruited from Columbia University Irving Medical Center, with the same inclusion criteria as study 1. In an initial visit, participants received a one-hour nutrition training introducing relevant concepts like food “choices” and macronutrients. An investigator introduced participants to the GlucoGoalie application, helped them set an initial goal of their choice, and asked them to log meals at home with BG readings before and two-hours after meals, for 4 weeks. During training, we told participants that GlucoGoalie would recommend

Table 3: Prevalence of themes across participant interviews. Each purple-shaded cell indicates that a theme was present for a participant. Theme 1 was prevalent in 100% of interviews, while Theme 4 was prevalent in 50% of interviews.

	P1	P2	P3	P4	P5	P6	P7	P8
Theme 1 - Receiving goal suggestion informs self-discovery	Present	Present	Present	Present	Present	Present	Present	Present
Theme 2 - Choosing goals highlights individual preferences	Present	Present	Present	Present	Present	Present	Present	Present
A checkpoint or a challenge	Present	Present	Present	Present	Present	Present	Present	Present
Importance of personal food preferences	Present	Present	Present	Present	Present	Present	Present	Present
Theme 3 - Following goals demonstrates the importance of feedback and context	Present	Present	Present	Present	Present	Present	Present	Present
Fitting goals with the context of daily life	Present	Present	Present	Present	Present	Present	Present	Present
Importance of feedback and seeing progress	Present	Present	Present	Present	Present	Present	Present	Present
Theme 4 - Challenges understanding and following goals in practice	Present	Present	Present	Present	Present	Present	Present	Present
Balancing abstract and concrete in nutrition goals	Present	Present	Present	Present	Present	Present	Present	Present
Imprecision of text for delivery of goal suggestions	Present	Present	Present	Present	Present	Present	Present	Present

goals based on their own records, that these goals were made by a computer, not a human expert, and that available goals would change over time. In addition to push notifications, the study coordinator also contacted participants the very first time they received personalized goals. After the 4-week period, participants returned for 1-hour semi-structured interviews. To minimize barriers to participation, individuals without smartphone received an Android phone and could keep it after completing the study (participants who had their own smartphones received its monetary equivalent, \$150). All participants received \$20 for the initial visit and debrief interview.

5.1.2 Data Analysis. We downloaded usage log data from the application server and calculated descriptive usage statistics including the numbers of meals logged, goals selected, and goals used. Debrief interviews were audio recorded and transcribed verbatim. We analyzed interview transcripts and usage logs with inductive thematic analysis [12]. The lead author coded 2 transcripts (25%) collaboratively with a second author to create an initial codebook. Then the first and senior author independently coded an additional 2 transcripts (25%), and met in person to discuss coding schemes and resolve all discrepancies through discussion. The remaining interviews were coded independently by the first author with periodic discussion with the research team, followed by an affinity mapping session to group codes into primary themes and subthemes. Participant meal logs, usage, and goal attainment were considered throughout the coding process to contextualize user statements. After coding was complete, we examined data saturation and theme comprehensiveness across participants [39, 43].

5.2 Deployment Study Results

Below, we briefly describe participants' demographics and usage of the GlucoGoalie app, and then continue to describe the four main themes from the thematic analysis: 1) receiving goal suggestion informs self-discovery, 2) choosing goals highlights individual preferences, 3) following goals demonstrates the importance of feedback and context, and 4) challenges understanding and following goals in practice. As shown in Table 3, data saturation was reached and themes were prevalent across participants.

5.2.1 Participant Background and Demographics. A total of 10 participants were enrolled in the deployment study. Two participants were lost to follow-up or withdrew for personal reasons, for a total of 8 participants who completed the deployment study and were included in the analysis. There was some overlap between the controlled experiment and the deployment study, with 2 participants completing both studies. As shown in Table 2, the demographic breakdown was comparable across the two studies. While median income was higher in the deployment study, the distributions were similar, and the deployment study clinics were in an area with a higher cost of living (Manhattan vs. Brooklyn).

Participants had mixed and often poor experiences with self-management prior to enrolling in the study. Many reported poor eating habits and being indiscriminate about their meals:

“... before that I eat whatever. Yeah, whatever. Dinner time, I eat whatever.” P2

Others often skipped meals, which led to overeating later in the day:

“... I only skip breakfast. I wasn't always very good about lunch. So, then I'll be famished. I would eat crap because I was hungry.” P5

Along with challenges with nutrition, participants described challenges keeping their BG within target ranges:

“Sometimes [my blood sugar] goes very high or goes very low.” P4

“Yeah, a mess, my sugar level was high everyday 300, and the doctor was upset to me.” P1

Some participants had tried prior bouts of focused self-management, with mixed success in the long run. Three participants had previously tracked their meals on paper, but none had tracked with an app.

5.2.2 Impressions and Usage. Overall, participants reported that they enjoyed the experience of using GlucoGoalie, and found it fun, easy, and direct.

“It was fun. They laugh about me because every time I was going to eat — *no, wait a minute. I can't start eating. I've got to take a picture of it.* ... It was fun to play.” P2

Table 4: GlucoGoalie usage statistics over the 4-week study (N=8)

Usage Variable	Value
# meals logged	Median: 93.5 Range: 14 to 158
# BG readings logged	Median: 173 Range: 21 to 314
% received personalized goals	88% (7 of 8)
% used personalized goals	63% (5 of 8)
# unique personalized goals used during study period	Median: 3 goals Range: 0 to 8
Pre-meal BG (mean ± SD)	135.0 ± 36.7 mg/dL
Post-meal BG (mean ± SD)	156.1 ± 47.7 mg/dL

They also actively engaged with the main part of the app: setting and following goals.

“I try to follow the goals and instructions if I’m trying to improve my intake. That’s what I’m trying to do most of the time. Because every day I try to follow a better diet and try to have more greens.” P4

Usage statistics over the 4-week study are presented in Table 4. Participants showed high engagement with logging features: on average they recorded more than 3 meals per day, and all participants set at least one goal. However, only about 40% (3 of 8) of participants actively engaged with different features of the app, such as setting new goals and viewing progress towards goal achievement; these were savvy users of smartphones with previous experience using apps. In contrast, most of the participants in the study (5 of 8; about 60%) were more accustomed to using their smartphones exclusively to make phone calls and rarely used any apps. These individuals often took a minimalist approach to engaging with GlucoGoalie: they tracked their meals and assessed these meals on fit with chosen goals, but did not engage with any other features without prompts from investigators.

Regarding personalized goals, 88% of participants (7 of 8) received personalized goals while in the study; one participant did not receive any personalized goals because their BG levels were well-controlled. Of those who received a goal recommendation, 71% (5 of 7) selected at least one of these goals in the app. However, 3 participants did not notice a push notification informing them of a new goal suggestion, and only selected one after a call from the study coordinator.

As a result of following the goals they had chosen, many participants developed new habits, and internalized personalized goal suggestions to the point that they became integrated with their daily practice:

“Even anything longer than two weeks will probably just make it into more of a habit for me. I’ll probably eat two weeks to get comfortable with how much fat I’m taking, let’s say the goal was on fat, so then after that it would just be more of a habit.” P8

At the end of the study, many participants described seeing changes not just in their behaviors, but also in their actual blood glucose levels.

“I did notice because sometimes it was 200. When I see that it was 200, it was after I eat. Oh yeah. After I—but before, 250, 270—because I was eating a lot of food. Five or six in the night.” P2

“And the sugar went down. . . . Today, I tested, it was 121.” P6

5.3 Theme 1—Receiving Goal Suggestion Informs Self-Discovery

To personalize goals, GlucoGoalie included features for tracking meals and BG levels. The study showed that even these requisite tracking features often led to discoveries and new insights. Furthermore, the experience of viewing both generic and personal goal suggestions helped individuals critically reflect on their behaviors, thus serving as an additional catalyst for learning.

Through tracking and reflecting on their meals, participants described some of the patterns and insights they observed between the foods they were eating and their BG levels.

“I did it for two days and I tested my sugar, *oh, this is the rice*. . . . So, I stopped eating rice for two days, and then when I stopped eating rice, it got lower.” P6

Beyond tracking, the goal setting features in GlucoGoalie scaffolded the self-discovery process. For example, P2 learned from their goal to “eat whole fruits instead of juice”

“When I drink the juice, I see that sugar is what was high. And I learned that that was the problem. . . . Now, when I eat, I don’t drink juice.” P2

In many cases, participants used the personalized goal recommendations they received to reflect on their behaviors and sought to reconcile these recommendations with what they had already knew or suspected about themselves.

“And I know that, my carbs like I said, are usually high. I think that, my first, what I gravitate to first in any meal is the carb and that’s what I want more of. . . . So, I’m not like surprised that it recommends reducing the carbs and trying to replace it with something else.” P3

Participants sometimes noted that the goal they received was something they were already trying to work on. For example, P4 described their reaction to receiving a suggestion to reduce the amount of fat in their meals:

“I’m trying to decrease the amount of food and so that’s why, I think it’s important to decrease the amount of fat and that is one of the problems that I have with the fat.” P4

Receiving personalized goal suggestions provided a reference point for participant’s own views of their self-management pitfalls and needs, as well as a jumping off point to guide reflection on their behaviors.

5.4 Theme 2—Choosing Goals Highlights Individual Preferences

5.4.1 A Checkpoint or a Challenge. Most participants commented that some goals in the GlucoGoalie app seemed harder to achieve than others. However, when choosing which goals to follow, participants took a variety of different approaches. Some participants chose goals that seemed highly achievable, or were the sorts of behaviors they were already doing regularly; these participants viewed goals as a checkpoints or reminders to be more consistent.

“I like that it was a goal that it was more feasible to me. So, it was just a good like a checkpoint for me not sort of a reminder but kind of like, oh it’s going with what I’m doing. So, it’s just reminding me.” P8

In contrast, other participants were interested in choosing goals that were more challenging as self-motivation to change their current habits.

“Yes, I go to the notification and started looking at the new one. That’s why, when I first took the other substitute of water for over sodas. I realized, well that’s not really a goal because I’ve been doing that already. So, I need to change to something more difficult because I was done with the other one.” P4

5.4.2 Importance of personal food preferences. In addition to the perceived degree of challenge in a given goal, personal likes and dislikes regarding different foods factored in prominently to participants’ decisions of which goal to choose. To illustrate personalized goals, GlucoGoalie included three examples with different foods at the end of each goal (see Table 1). For many participants, these examples were critical factors to deciding whether to try a goal or not. When asked to explain why they selected a particular personalized goal, participants often referenced the examples as their justification for selecting or eliminating a goal from consideration.

“That one is okay, because I used to eat the oatmeal, one slice of toast, yeah that one is okay.” P1

Along with expressing their interest or distaste for certain foods, participants also mentioned the importance of variety, and opted for suggestions that incorporated new ideas to break what they perceived as the monotony of healthy eating. For example, P2 was looking for examples of vegetables they could eat other than broccoli:

“I don’t know, like, if I want to eat like broccoli, I will be tired. And I’m not going to eat it every day.” P2

5.5 Theme 3—Following Goals Demonstrates the Importance of Feedback and Context

5.5.1 Fitting Goals with the Context of Daily Life. The need for greater personalization extended beyond choosing which goal to pursue and impacted participants’ ability to successfully incorporate new goals within their daily lives. In some cases, participants had established patterns that they did not want to change, for example eating the same thing for breakfast every day because it worked for them, or skipping breakfast entirely because their morning routine did not allow for it. Furthermore, balancing meals within a day

or week was just as important. What made sense for an upcoming meal depended in part on what happened earlier in the day.

“Since I’m a busy woman. . . it kind of just has to go back to like how my day is. So, I know that if I didn’t meet it for one of my meals, I’ll have to meet it for the next meal.” P8

This balance extended to seasonal patterns as well, where different kinds of meals were appealing during different parts of the year.

“I don’t want to have a hearty breakfast compared to like in the winter.” P8

Many participants touted that it was easier to follow goals when preparing their own meals at home, but much harder when eating outside, at a restaurant or other gathering. Goals in GlucoGoalie lent themselves particularly to the home context, but different goals may be useful in other contexts.

“...well at least for me... it was very hard for me to manage using the app when I went out to eat.” P3

When goals felt appropriate also depended on the context of other self-management and health goals, for example exercise. P5 noted that they often include more carbs in their meals after exercising, but less if they have not exercised that day:

“So, I know, if I have exercise, walking or an exercise routine after a meal that’s going to be a little bit more high carbs. That has made an impact.” P5

5.5.2 Importance of Feedback and Seeing Progress. Participants were eager for feedback on their progress. This included whether they were successfully meeting the goals they had set in GlucoGoalie, for example, whether the amounts of specific macronutrients in their meals were more consistent with their chosen goals. Most participants found this challenging and had to come up with strategies. Some started measuring their foods to get a better sense for portion sizes and proportions:

“When I got after I started, I look for a [measuring] cup and I started to follow the instructions.” P4

In general, participants were eager for feedback on their progress:

“Everybody would like to know how they’re doing. . . Because if I’m eating less and it’s not doing no good, what’s the point of me doing it?” P2

In particular, many participants described not only the goals they had set with GlucoGoalie, but also their higher-level goals, motivations, and aspirations. These goals were not at the specific and achievable level of “drink more water,” but reflected general desires for leading a healthy life. Importantly, different participants expressed different motivations. Some participants expressed a desire to lose weight, or to see that their blood glucose levels were lowering.

“Definitely in terms of weight loss but like also my actual numbers in terms of my blood sugar.” P3

Other participants were also interested in improving their diabetes management, and had the goal of improving control of blood glucose levels, so that they could reduce their dosage of oral medications like metformin.

“Because I want to keep it as level as possible to try to stay off medications.” P7

5.6 Theme 4—Challenges Understanding and Following Goals in Practice

5.6.1 Balancing Abstract and Concrete in Nutritional Goals. Nutritional goals in GlucoGoalie included references to both specific foods and food groups, such as “Drink more water” and also macronutrients, such as “replace 1 carb choice with 1 protein choice at lunch”. Many participants’ comments related to the interplay between abstract and concrete when thinking about nutrition.

In general, participants enjoyed goals that were concrete and easy to implement without additional knowledge. This was particularly the case for generic goals that typically targeted familiar foods or food groups.

“Those were right. Those were easy and I’ve been, I have been intentional to drink a bottle of water at every main meal and then have a bottle or two in between.” P5

However, personalized goals were more abstract with a focus on macronutrients rather than specific foods. These goals were typically described as harder to understand and meet.

“The replacement, it was, you know was dropping, half a carb replacing, half carb. That was a little harder to figure out. So, it will require a little more thinking.” P5

Furthermore, participants’ attitudes towards more abstract, macronutrient-oriented goals were influenced by their apparent knowledge of nutrition. About half of participants were comfortable identifying macronutrients, estimating portion size, and discussing steps they could take to meet these goals with their meal choices.

“So, I still go by the basics even from when I went to the nutritionist of like using like my palms, like the two fingers, index fingers. Actually, do work well for like teaspoons and tablespoons.” P8

The other half of participants described themselves as not being familiar with macronutrients and estimating portion sizes. For these participants, goals formulated using macronutrients and “choices” as units presented an impassable barrier and were often dismissed. These participants often referred to using visual proportions of different types of foods on their plate to gauge how healthy their meals were:

“I use my plate, but I try to go as they show me in the program, you see the plate then half it’s a vegetable or fruit, this is a protein and that one is a carbohydrate.” P1

5.6.2 Imprecision of Text for Delivery of Goal Suggestions. Even for those with higher nutrition literacy, participants were not always consistent in how they interpreted personalized goals, and there were a number of misunderstandings. For example, some terminology, like “choices” as a unit of measure, was often interpreted as an option to choose two different food items, regardless of the amount. P2 described their effort to achieve a goal of eating 2 fat

choice (10g) at breakfast by stating that they ate two high fat food, but not the amounts of either:

“Sometimes I put it together, the mozzarella on top of the egg which means I’m taking two fats.” P2

While this meal may have been consistent with P2’s goal, they are saying they believe they achieved their goal because they chose two fat-based ingredients, not because the amount of total fat in the meal is consistent with the goal.

In addition, participants sometimes struggled with the numerical content in goals, for example the combination of both “choices” and “grams” as units.

“‘Decrease your fat to about four fat choices.’ That part is pretty clear. The only part that I say, kind of gets tricky where I guess you’re adding numbers with words would be the ‘20 grams.’” P8

In general, static text alone was limited in its ability to convey the more abstract nutrition goals. During the interviews, participants asked a number of clarifying questions, for example asking which foods count as which macronutrients. Some participants suggested that visual aids for portion size estimation would be a welcome addition.

6 DISCUSSION

The goal of this research was to examine individuals’ experiences with receiving, selecting, and following computationally generated nutritional goals for T2D. In designing GlucoGoalie, we took the approach of combining ML analysis of individuals’ self-tracking data with an expert system to computationally generate recommendations for nutritional goals that are likely to lead to improvement in BG levels.

This approach has several important distinctions compared to previously proposed systems. First, the ML inference in GlucoGoalie directly examines the relationship between behavior and a health marker (BG) to inform recommendations; not by assuming which behaviors are healthy [83], or relying on user’s self-perceptions of what behaviors impact health [47]. GlucoGoalie makes recommendations in the multidimensional space of nutritional composition, versus the unidimensional space of steps [58] or calories [83], which makes it more complex. Furthermore, unlike other recommendation approaches (e.g., MyBehavior [83]), integration of expert knowledge within the expert system enables GlucoGoalie to make suggestions that extend beyond individuals’ past behaviors (previously captured meals).

We evaluated this approach in two studies that examined whether individuals can understand and follow nutritional goals generated by the expert system using ML inferences, and what experiences would result from receiving and following such goals in the context of individuals’ daily lives.

In the controlled experiment, we found that participants were largely able to understand and act on computationally derived goals in a controlled setting. Participants correctly chose meals that met a goal 89% of the time when these meals were accompanied by corresponding nutritional labels. When composing meals to meet a chosen goal at a “virtual buffet,” participants assembled meals in the correct direction of the goal 67% of the time. This suggests that individuals were able to understand the personalized goals, and

were moderately successful when composing meals to meet goals. At the same time, additional findings highlight the complexity of nutrition decisions. When choosing which of two meal images met a given goal without nutrition labels, participants were correct only 49% of the time. This aligns with prior research suggesting that individuals have difficulty comparing macronutrient quantities from photographs alone [13]. In addition, participant comments during the study indicated confusion about some of the nutrition terminology in goals, and there was considerable variability in meeting the exact goal target. This suggests that participants formed a general idea of how to achieve goals, but had difficulty precisely implementing the recommendations.

The deployment study found similar successes and challenges for participants in understanding and acting on goal recommendations. Specifically, it showed that participants were generally able to understand goals, and at least attempted to follow them. The results also reiterated challenges related to specific design choices, like the use of the word “choice” to describe macronutrient quantities.

In addition, the deployment study revealed a number of insights related to the experience of receiving and following goal suggestions in everyday life. Specifically, it highlighted the relationship between supporting reflection and direct support for action, the alignment between goals with individuals eating practices and larger aspirations, and the need for interactive approaches that enable feedback and negotiation. Below we discuss these themes and relate them to our design choices.

6.1 Balancing Support for Reflection and Action

Personal informatics aims to increase self-knowledge and, ultimately, inform future action through collection of and reflection on self-tracking data [64]. However, reflecting on data can be burdensome, and not everyone has the necessary time, mental energy, and literacy. In contrast, there is a long tradition of research in behavior change interventions that focus less on reflection and provide more direct support for action through a variety of behavior change techniques [72]. One limitation of traditional behavior change interventions is that they rely on predetermined behavior goals to nudge users towards, but in the case of chronic conditions like T2D, different goals may be appropriate for different individuals based on their physiology and response to diet. While a more direct approach may mitigate the burden of reflection, a potential concern is that it could lead to individuals following the system’s recommendations without attaining the benefits of learning and self-discovery, which could have a negative impact on autonomy [49].

Our study suggested that it is possible to reach a middle ground between these extremes. Because GlucoGoalie used an expert system to generate concrete goal recommendations, it was able to provide direct support for action. At the same time, because goals were informed by ML analysis of self-tracking data, the participants often engaged in reflection similar to the one enabled by personal informatics solutions. The participants appreciated the more direct support for action through goal recommendations: those who selected personalized goals in the app described making changes and choosing meals that would be consistent with goals, for example taking increased care to measure the components of their meal. At

the same time, the study showed that participants found tracking meals and BG levels to be informative, an experience similar to most personal informatics solutions [47, 65]. Furthermore, we found that participants actively engaged with the recommendations they received and took them as an additional prompt and opportunity for reflection, beyond that provided by the personal data itself. Participants compared goal suggestions to their own self-perceptions of their eating habits and used them as a mirror to re-examine their past choices. In this way, we found a synergy between offering direct support for action as a part of an application that enables reflection via self-tracking.

These findings highlight the potential for solutions that balance support for both reflection and action. First, future work could more directly explore the relationship between actionable recommendations and reflection in self-tracking, for example comparing engagement in self-tracking with and without the addition of actionable recommendations. Second, in this work, the connection between one’s behaviors and the recommendations they received were not explained or made explicit by the application, but relied on users to fill in those gaps. Future work could endeavor to make the connections between personal data and recommendations more salient for users, which may further support engagement and reflection. For example, actionable recommendations could be enhanced by presenting visual summaries of the self-tracking data that informed the specific goal recommendations [34, 91]. This additional information can serve as a form of explanation for the recommendations, and prior work has demonstrated the importance of explanations in facilitating nutritional learning [13]. A growing body of research in explainable ML may offer potential avenues to make recommendations in support of action and ground them with an explanation to support reflection [100]. Future work could further incorporate advances in explainable ML to personal informatics applications.

6.2 Aligning Goals with Eating Experiences and Personal Aspirations

Because GlucoGoalie relied on an expert system to generate recommendations as natural language sentences, one of our challenges was to find the right form to formulate these recommendations. Through the design process, we took the approach of formulating goals in terms of macronutrient amounts [36], which has the advantage of allowing individuals to flexibly apply their goal to different types of foods and meals, with the ability to freely incorporate their food preferences. However, this study demonstrated some limitations of this approach. While participants who expressed comfort with nutrition terminology were able to adopt goals, those with lower nutrition literacy and less comfort measuring or weighing their food had trouble understanding and following goals. Making meal choices ultimately comes down to what’s on one’s plate, and participants sometimes found it difficult to connect somewhat abstract goals to concrete meal choices.

An alternative and common form of nutrition suggestions are recipe or meal plan recommendations, which are much more concrete and consistent with how participants think about their meals and diet. However, as recommendations become more concrete, they need to take into account individual’s food preferences, and there are more opportunities to miss the mark. We found this with

the “examples” included with each personalized goal: idiosyncratic preferences for a single food item in the list was a major factor in whether a participant would choose a goal or not. Recommender systems (RecSys) excel at making concrete suggestions based on personal preferences, learned from users’ past behavior or characteristics [87], and can incorporate additional constraints like food allergies [48]. While GlucoGoalie focused on personalizing recommendations based on health constraints, this approach could be complimentary with growing research in health-aware RecSys [32]. Meal logs and macronutrient-centered goals from GlucoGoalie could be used as inputs to a preference-based RecSys to generate concrete suggestions that would help individuals connect their goals to what’s on their plate.

In addition to food preferences, participants highlighted the importance of context in determining when a goal was appropriate, for example the time of year, how active one has been, and what other meals have been eaten recently. Making contextually-appropriate recommendations adds another dimension of complexity [84]. Mobile phones and sensors can offer clues to the user’s current state, and there is a long history of research in context-aware computing within HCI and Ubiquitous Computing [28]. In health, location-based prompts have been used to help prevent relapse triggers [17], and step counts can inform adaptive fitness goals based on recent activity levels [58], but have not been widely used in nutrition [83].

A final tension was participants’ desire for a greater connection between specific nutritional goals and their larger aspirations in life and health. Participants did not always see the connection between concrete, quantifiable self-tracking-related goals and larger, more abstract, qualitative motivations. Niess and Woźniak observed the relationship between tracking goals and qualitative health goals in the context of individuals setting goals with fitness trackers [77]. For example, a quantitative, self-tracking goal of walking 12k steps a day might be connected to a qualitative goal of losing weight, and a higher-level goal of feeling well. In the case of GlucoGoalie, because the algorithm suggests quantitative goals, it’s even more important to draw a connection back to an individual’s qualitative goals, like improving BG levels. Researchers have explored methods to elicit these values and motivations [7] and future work could explore how to connect them to quantitative tracking goal [77].

6.3 Interactivity, Negotiation, and Feedback

By taking the approach of using an expert system to interpret ML output, GlucoGoalie produced static, text-based recommendations. One of the limitations of this approach was that we were unable to resolve the misinterpretations and misunderstandings that are likely to arise in a complex domain like nutrition. In some cases, participants did not understand the nutrition terminology, and in other cases they understood the vocabulary, but misinterpreted the intended meaning. One approach to make nutrition goals more understandable is to incorporate illustrations. In health risk communication, illustrations and infographics have been used successfully to improve comprehension of complex information [3, 42, 104]. A similar visual approach has been applied to assist low literacy adults with portion size estimation [16], and could be used here to better convey numerical content in personalized goals.

A second approach is to offer the opportunity for questions and answers in a back-and-forth exchange. This more interactive

approach could introduce concepts, answer users’ questions, and more fully explain goal recommendations. Along these lines, conversational agents have been used to support interactive goal setting, health coaching, and motivational interviewing [10, 62]. Generally, these approaches are based on a set list of goals, not personalized based on user self-tracking data. Combining conversational agents with computationally personalized goal setting is a potential direction for future work. A more interactive and conversational interaction style would also offer another approach to address the challenges of context, discussed above, to allow participants to have input on their goals and negotiate [56]. Finally, this approach might also address the lack of proactive engagement from some users, particularly those with less technology comfort, who did not explore app features and sometimes did not notice updates to their available goals. While many smartphone features rely on users accessing features to pull information or support, conversational approaches can proactively initiate interactions, which may lead to a higher level of engagement with these features [8, 38, 93]. Enabling negotiation within the space of possible goals expands on the complexity of recommendations, and may require more sophisticated and flexible methods than the rule-based expert system used in GlucoGoalie. Machine learning approaches like mechanistic, controller, or reinforcement learning models are a potential vein of future exploration [1, 63, 69].

In another opportunity for increased interactivity, participants expressed resounding interest in more feedback about their progress in achieving their goals, and the impact of this progress on their overall health. Feedback is an important component of learning in goal-setting [29], and while participants were able to self-assess each meal against their goal and view a summary of their goal attainment, they were interested in additional feedback from GlucoGoalie. One approach to providing feedback is to engage dietitians and other healthcare professionals. However, this increases reliance on human experts, thus limiting the scalability of the approach. Previous research in coaching interventions explored offering automated feedback, especially for physical activity [20, 45, 83]. Similar techniques could be applied to nutrition in future research.

6.4 Limitations

As an initial step towards exploring actionable health recommendations, this work has a number of limitations. Both studies had small sample sizes, and while the sample was recruited from economically disadvantaged communities, it was not representative: participants were skewed female, and predominantly black or Latino. This cohort from a single United States metro area, may not account for important cultural differences nationally or globally [94]. In addition, the study ran for 4-weeks, and usage patterns and engagement may change with extended use. Finally, while we report qualitative perceptions, we did not quantitatively examine changes in BG or other health outcomes because of the small sample and short timeframe [54].

6.5 Conclusion

While self-tracking data hold potential to inform action, deriving actionable insights is challenging and burdensome for individuals. In this work, we explored an approach that combines ML with an

expert system to generate goals that are personalized based on an individual's health data. We found that support for action can also support, augment, and inform reflection. We connected our findings with both prior research and open questions in personal informatics, goal-setting, and health-based recommender systems, and argue that future interventions could incorporate more interactive, dialog-based components with conversational agents.

ACKNOWLEDGMENTS

This research was funded by the National Institute of Diabetes and Digestive and Kidney Diseases award number R56DK113189 and the National Library of Medicine award number T15LM007079. Thank you to the fellow students of Columbia's Department of Biomedical Informatics, who earned the undying gratitude of the corresponding author by arts-and-crafting food images for the virtual buffet.

REFERENCES

- David J Albers, Matthew E Levine, Andrew Stuart, Lena Mamykina, Bruce Gluckman, and George Hripcsak. 2018. Mechanistic machine learning: how data assimilation leverages physiologic knowledge using Bayesian inference to forecast the future, infer the present, and phenotype. *Journal of the American Medical Informatics Association* 25, 10: 1392–1401. <https://doi.org/10.1093/jamia/ocy106>
- American Diabetes Association. 2018. 4. Lifestyle Management: Standards of Medical Care in Diabetes-2018. *Diabetes care* 41, Suppl 1: S38–S50. <https://doi.org/10.2337/dc18-S004>
- Adriana Arcia, Niurka Suero-Tejada, Michael E. Bales, Jacqueline A. Merrill, Sun-moo Yoon, Janet Woollen, and Suzanne Bakken. 2016. Sometimes more is more: Iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *Journal of the American Medical Informatics Association* 23, 1: 174–183. <https://doi.org/10.1093/jamia/ocv079>
- Pablo Aschner. 2017. New IDF clinical practice recommendations for managing type 2 diabetes in primary care. *Diabetes Research and Clinical Practice* 132: 169–170. <https://doi.org/10.1016/j.diabres.2017.09.002>
- Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing reflection. In *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*, 93–102. <https://doi.org/10.1145/2598510.2598598>
- Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Transactions on Computer-Human Interaction* 20, 5: 1–27. <https://doi.org/10.1145/2503823>
- Andrew B.L. Berry, Catherine Lim, Andrea L. Hartzler, Tad Hirsch, Evette Ludman, Edward H. Wagner, and James D. Ralston. 2017. Eliciting Values of Patients with Multiple Chronic Conditions: Evaluation of a Patient-centered Framework. *AMIA ... Annual Symposium proceedings. AMIA Symposium 2017*: 430–439. Retrieved September 15, 2020 from <https://pmc/articles/PMC5977727/?report=abstract>
- Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59, 1: 21–30. <https://doi.org/10.1016/j.pec.2004.09.008>
- Timothy W. Bickmore, Rebecca A. Silliman, Kerrie Nelson, Debbie M. Cheng, Michael Winter, Lori Henault, and Michael K. Paasche-Orlow. 2013. A Randomized Controlled Trial of an Automated Exercise Coach for Older Adults. *Journal of the American Geriatrics Society* 61, 10: 1676–1683. <https://doi.org/10.1111/jgs.12449>
- Timothy W Bickmore, Daniel Schulman, and Candace L Sidner. 2011. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of Biomedical Informatics* 44, 2: 183–197. <https://doi.org/10.1016/j.jbi.2010.12.006>
- Thomas Bodenheimer, Kate Lorig, Halsted Holman, and Kevin Grumbach. 2002. Patient Self-management of Chronic Disease in Primary Care. *JAMA* 288, 19: 2469. <https://doi.org/10.1001/jama.288.19.2469>
- Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2: 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Marissa Burgermaster, Krzysztof Z. Gajos, Patricia Davidson, and Lena Mamykina. 2017. The Role of Explanations in Casual Observational Learning about Nutrition. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4097–4145. <https://doi.org/10.1145/3025453.3025874>
- Kerri L Cavanaugh. 2011. Health literacy in diabetes care: explanation, evidence and equipment. *Diabetes management (London, England)* 1, 2: 191–199. <https://doi.org/10.2217/dmt.11.5>
- Centers for Disease Control and Prevention. 2018. Racial and Ethnic Approaches to Community Health | DNPAO. Retrieved January 3, 2019 from <https://www.cdc.gov/nccdphp/dnpao/state-local-programs/reach/>
- Beenish M. Chaudhry, Christopher Schaeferbauer, Ben Jelen, Katie A. Siek, and Kay Connelly. 2016. Evaluation of a Food Portion Size Estimation Interface for a Varying Literacy Population. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 5645–5657. <https://doi.org/10.1145/2858036.2858554>
- Ming-Yuan Chih, Timothy Patton, Fiona M. McTavish, Andrew J. Isham, Chris L. Judkins-Fisher, Amy K. Atwood, and David H. Gustafson. 2014. Predictive modeling of addiction lapses in a mobile health application. *Journal of Substance Abuse Treatment* 46, 1: 29–35. <https://doi.org/10.1016/j.jsat.2013.08.004>
- 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 32nd Annual ACM Conference on Human Factors in Computing Systems*: 1143–1152. <https://doi.org/10.1145/2556288.2557372>
- 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* 3, 1: 1–27. <https://doi.org/10.1145/3314394>
- Céline Clavel, Steve Whittaker, Anaïs Ana is Blacodon, and Jean-Claude Martin. 2018. WEnner: A Theoretically Motivated Approach for Tailored Coaching About Physical Activity. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp '18 (UbiComp '18)*, 1669–1675. <https://doi.org/10.1145/3267305.3274190>
- James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, 647–658. <https://doi.org/10.1145/2750858.2807554>
- Heather J. Cole-Lewis, Arlene M. Smaldone, Patricia R. Davidson, Rita Kukafka, Jonathan N. Tobin, Andrea Cassells, Elizabeth D. Mynatt, George Hripcsak, and Lena Mamykina. 2016. Participatory approach to the development of a knowledge base for problem-solving in diabetes self-management. *International Journal of Medical Informatics* 85, 1: 96–103. <https://doi.org/10.1016/j.ijmedinf.2015.08.003>
- Francis S. Collins and Harold Varmus. 2015. A New Initiative on Precision Medicine. *New England Journal of Medicine* 372, 9: 793–795. <https://doi.org/10.1056/NEJMp1500523>
- Sunny Consolvo, Predrag Klasnja, David W. McDonald, and James A. Landay. 2009. Goal-setting considerations for persuasive technologies that encourage physical activity. In *ACM International Conference Proceeding Series*, 1. <https://doi.org/10.1145/1541948.1541960>
- Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the Mobile Food Journal: Exploring Opportunities for Lightweight Photo-Based Capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3207–3216. <https://doi.org/10.1145/2702123.2702154>
- Nediyana Daskalova, Jina Yoon, Yibing Wang, Cintia Araujo, Guillermo Beltran, Nicole Nugent, John McGeary, Joseph Jay Williams, and Jeff Huang. 2020. Sleep-Bandits: Guided Flexible Self-Experiments for Sleep. In *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3313831.3376584>
- 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. In *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3290605.3300600>
- Anind K. Dey. 2001. Understanding and using context. *Personal and Ubiquitous Computing* 5, 1: 4–7. <https://doi.org/10.1007/s007790170019>
- The Diabetes Prevention Program (DPP) Research Diabetes Prevention Program (DPP) Research Group. 2002. The Diabetes Prevention Program (DPP): description of lifestyle intervention. *Diabetes care* 25, 12: 2165–71. <https://doi.org/10.2337/diacare.25.12.2165>
- Sara Belle Donevant, Robin Dawson Estrada, Joan Marie Culley, Brian Habing, and Swann Arp Adams. 2018. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. *Journal of the American Medical Informatics Association*. <https://doi.org/10.1093/jamia/ocy104>
- David Elswailer and Morgan Harvey. 2015. Towards Automatic Meal Plan Recommendations for Balanced Nutrition. *Proceedings of the 9th ACM Conference on Recommender Systems*: 313–316. <https://doi.org/10.1145/2792838.2799665>
- David Elswailer, Bernd Ludwig, Alan Said, Hanna Schaefer, and Christoph Trattner. 2016. Engendering Health with Recommender Systems. In *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, 409–410. <https://doi.org/10.1145/2959100.2959203>

- [33] 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 2016 CHI Conference on Human Factors in Computing Systems*: 5632–5644. <https://doi.org/10.1145/2858036.2858044>
- [34] Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. *DIS '14 Proceedings of the 2014 conference on Designing interactive systems*: 667–676. <https://doi.org/10.1145/2598510.2598558>
- [35] Deborah Estrin. 2014. Small data, where n = me. *Communications of the ACM* 57, 4: 32–34. <https://doi.org/10.1145/2580944>
- [36] Alison B. Evert, Michelle Dennison, Christopher D. Gardner, W. Timothy Garvey, Ka Hei Karen Lau, Janice MacLeod, Joanna Mitri, Raquel F. Pereira, Kelly Rawlings, Shamera Robinson, Laura Saslow, Sacha Uelmen, Patricia B. Urbanski, and William S. Yancy. 2019. *Nutrition therapy for adults with diabetes or prediabetes: A consensus report*. <https://doi.org/10.2337/dci19-0014>
- [37] Daniel J Feller, Marissa Burgermaster, Matthew E Levine, Arlene Smaldone, Patricia G Davidson, David J Albers, and Lena Mamykina. 2018. A visual analytics approach for pattern-recognition in patient-generated data. *Journal of the American Medical Informatics Association*. <https://doi.org/10.1093/jamia/ocy054>
- [38] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR mental health* 4, 2: e19. <https://doi.org/10.2196/mental.7785>
- [39] Patricia Fusch and Lawrence Ness. 2015. Are We There Yet? Data Saturation in Qualitative Research. *The Qualitative Report* 20, 9. Retrieved September 15, 2020 from <https://nsuworks.nova.edu/tqr/vol20/iss9/3>
- [40] Andrea Grimes, Martin Bednar, Jay David Bolter, and Rebecca E. Grinter. 2008. EatWell: Sharing nutrition-related memories in a low-income community. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 87–96. <https://doi.org/10.1145/1460563.1460579>
- [41] Andrea Grimes, Vasudhara Kantroo, and Rebecca E. Grinter. 2010. Let's play! Mobile health games for adults. In *UbiComp '10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing*, 241–250. <https://doi.org/10.1145/1864349.1864370>
- [42] Lisa Grossman, Steven Feiner, Elliot Mitchell, and Ruth Masterson Creber. 2018. Leveraging Patient-Reported Outcomes Using Data Visualization. *Applied Clinical Informatics* 09, 03: 565–575. <https://doi.org/10.1055/s-0038-1667041>
- [43] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How Many Interviews Are Enough? *Field Methods* 18, 1: 59–82. <https://doi.org/10.1177/1525822X05279903>
- [44] Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA - Journal of the American Medical Association* 316, 22: 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- [45] Ankit Gupta, Tim Heng, Chris Shaw, Linda Li, and Lynne Feehan. 2018. Towards developing an e-coach to support arthritis patients in maintaining a physically active lifestyle. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth '18*, 392–395. <https://doi.org/10.1145/3240925.3240954>
- [46] Mark D. Hayward, Toni P. Miles, Eileen M. Crimmins, and Yu Yang. 2000. The Significance of Socioeconomic Status in Explaining the Racial Gap in Chronic Health Conditions. *American Sociological Review* 65, 6: 910. <https://doi.org/10.2307/2657519>
- [47] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. 2017. What Does All This Data Mean for My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. *Human-Computer Interaction* 32, 5–6: 208–267. <https://doi.org/10.1080/07370024.2016.1277724>
- [48] Paris Hsu, Jingshu Zhao, Kehan Liao, Tianyi Liu, and Chen Wang. 2017. AllergyBot: A Chatbot technology intervention for young adults with food allergies Dining out. In *Conference on Human Factors in Computing Systems - Proceedings*, 74–79. <https://doi.org/10.1145/3027063.3049270>
- [49] Bart A. Kamphorst. 2017. E-coaching systems: What they are, and what they aren't. *Personal and Ubiquitous Computing* 21, 4: 625–632. <https://doi.org/10.1007/s00779-017-1020-6>
- [50] Ravi Karkar, Jasmine Zia, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, and Roger Vilardaga. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 6850–6863. <https://doi.org/10.1145/3025453.3025480>
- [51] Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, and Julie A Kientz. 2016. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association* 23, 3: 440–448. <https://doi.org/10.1093/jamia/ocv150>
- [52] Shigeko Kato, Kayo Waki, Sadako Nakamura, Sanae Osada, Haruka Kobayashi, Hideo Fujita, Takashi Kadowaki, and Kazuhiko Ohe. 2016. Validating the use of photos to measure dietary intake: the method used by DialBetics, a smartphone-based self-management system for diabetes patients. *Diabetology International* 7, 3: 244–251. <https://doi.org/10.1007/s13340-015-0240-0>
- [53] Yoojung Kim, Sookyoung Ji, Hyunjeong Lee, Jeong-Whun Kim, Sooyoung Yoo, and Joongseok Lee. 2016. “My Doctor is Keeping an Eye on Me!”: Exploring the Clinical Applicability of a Mobile Food Logger. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 5620–5631. <https://doi.org/10.1145/2858036.2858145>
- [54] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In *Conference on Human Factors in Computing Systems - Proceedings*, 3063–3072. <https://doi.org/10.1145/1978942.1979396>
- [55] Predrag Klasnja and Wanda Pratt. 2012. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics* 45, 1: 184–198. <https://doi.org/10.1016/j.jbi.2011.08.017>
- [56] Rafal Kocielnik and Gary Hsieh. 2017. New Opportunities for Dialogue-based Interaction in Behavior Change Domain. In *CSCW 2017 workshop on Talking with Conversational Agents in Collaborative Action*. Retrieved October 30, 2019 from <https://talkingwithagents.files.wordpress.com/2017/02/7-kocielnik1.pdf>
- [57] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2: 1–26. <https://doi.org/10.1145/3214273>
- [58] Elizabeth V. Korinek, Sayali S. Phatak, Cesar A. Martin, Mohammad T. Freigoun, Daniel E. Rivera, Marc A. Adams, Pedja Klasnja, Matthew P. Buman, and Eric B. Hekler. 2018. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *Journal of Behavioral Medicine* 41, 1: 74–86. <https://doi.org/10.1007/s10865-017-9878-3>
- [59] Mark Kutner, Elizabeth Greenberg, and Justin Baer. 2006. A First Look at the Literacy of America's Adults in the 21st Century. NCES 2006-470. *National Center for Education Statistics*. Retrieved August 23, 2018 from <https://eric.ed.gov/?id=ED489066>
- [60] Liliana Laranjo, Annie Lau, and Enrico Coiera. 2017. Design and Implementation of Behavioral Informatics Interventions. Springer, Cham, 13–42. https://doi.org/10.1007/978-3-319-51732-2_2
- [61] Amanda Lazar, Christian Koehler, Joshua Tenenbaum, and David H. Nguyen. 2015. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, 635–646. <https://doi.org/10.1145/2750858.2804288>
- [62] Jisoo Lee, Eric B. Hekler, Emil Chiauzzi, Auriel Townner, and Marcy Fitz-Randolph. 2016. Helping Users Set Rules for Defining Short-Term Activity Goals. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, 2178–2184. <https://doi.org/10.1145/2851581.2892488>
- [63] Huitian Lei, Ambuj Tewari, and Susan A. Murphy. 2017. An Actor-Critic Contextual Bandit Algorithm for Personalized Mobile Health Interventions. Retrieved January 31, 2019 from <http://arxiv.org/abs/1706.09090>
- [64] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*: 557. <https://doi.org/10.1145/1753326.1753409>
- [65] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, 405. <https://doi.org/10.1145/2030112.2030166>
- [66] 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. <https://doi.org/10.1093/jamia/ocv187>
- [67] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and David Greenblatt. 2008. MAHI: Investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 477–486. <https://doi.org/10.1145/1357054.1357131>
- [68] Lena Mamykina, Drashko Nakikj, and Noemie Elhadad. 2015. Collective sense-making in online health forums. In *Conference on Human Factors in Computing Systems - Proceedings*, 3217–3226. <https://doi.org/10.1145/2702123.2702566>
- [69] César A. Martín, Daniel E. Rivera, Eric B. Hekler, William T. Riley, Matthew P. Buman, Marc A. Adams, and Alicia B. Magann. 2020. Development of a Control-Oriented Model of Social Cognitive Theory for Optimized mHealth Behavioral Interventions. *IEEE Transactions on Control Systems Technology* 28, 2: 331–346. <https://doi.org/10.1109/TCST.2018.2873538>
- [70] Nirupa R Matthan, Lynne M Ausman, Huicui Meng, Hocine Tighiouart, and Alice H Lichtenstein. 2016. Estimating the reliability of glycemic index values and potential sources of methodological and biological variability. *The American journal of clinical nutrition* 104, 4: 1004–1013. <https://doi.org/10.3945/ajcn.116.137208>

- [71] Mollie McKillop, Lena Mamykina, and Noémie Elhadad. 2018. Designing in the Dark. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–15. <https://doi.org/10.1145/3173574.3174139>
- [72] Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P. Eccles, James Cane, and Caroline E. Wood. 2013. The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. *Annals of Behavioral Medicine* 46, 1: 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- [73] Elliot G Mitchell, Esteban G Tabak, Matthew E Levine, Lena Mamykina, and David J Albers. 2021. Enabling personalized decision support with patient-generated data and attributable components. *Journal of Biomedical Informatics* 113, 103639: 103639. <https://doi.org/10.1016/j.jbi.2020.103639>
- [74] Sean Munson and Sunny Consolvo. 2012. Exploring Goal-setting, Rewards, Self-monitoring, and Sharing to Motivate Physical Activity. In *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*. <https://doi.org/10.4108/icst.pervasivehealth.2012.248691>
- [75] Aanand D. Naik, Nynikka Palmer, Nancy J. Petersen, Richard L. Street, Radha Rao, Maria Suarez-Almazor, and Paul Haidet. 2011. Comparative Effectiveness of Goal Setting in Diabetes Mellitus Group Clinics. *Archives of Internal Medicine* 171, 5: 453–459. <https://doi.org/10.1001/archinternmed.2011.70>
- [76] Brunilda Nazario. 2013. Portion Size Plate | Recommended Serving Sizes for Portion Control. Retrieved April 15, 2018 from <https://www.webmd.com/diet/healthtool-portion-size-plate>
- [77] Jasmin Niess and Paweł Wozniak. 2018. Supporting Meaningful Personal Fitness: the Tracker Goal Evolution Model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–12. <https://doi.org/10.1145/3173574.3173745>
- [78] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z. Gajos. 2011. Platemate: Crowdsourcing Nutritional Analysis from Food Photographs. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 1–12. <https://doi.org/10.1145/2047196.2047198>
- [79] Aisling Ann O’Kane, Sun Young Park, Helena Mentis, Ann Blandford, and Yunan Chen. 2016. Turning to Peers: Integrating Understanding of the Self, the Condition, and Others’ Experiences in Making Sense of Complex Chronic Conditions. *Computer Supported Cooperative Work: CSCW: An International Journal* 25, 6: 477–501. <https://doi.org/10.1007/s10606-016-9260-y>
- [80] Monica E Peek, Algernon Cargill, and Albert S Huang. 2007. Diabetes health disparities: a systematic review of health care interventions. *Medical care research and review: MCRR* 64, 5 Suppl: 101S–56S. <https://doi.org/10.1177/107758707305409>
- [81] Stephen Purpura, Victoria Schwanda, Kaiton Williams, William Stubler, and Phoebe Sengers. 2011. Fit4Life: The design of a persuasive technology promoting healthy behavior and ideal weight. In *Conference on Human Factors in Computing Systems - Proceedings*, 423–432. <https://doi.org/10.1145/1978942.1979003>
- [82] Aare Puussaar, Adrian K. Clear, and Peter Wright. 2017. Enhancing personal informatics through social sensemaking. In *Conference on Human Factors in Computing Systems - Proceedings*, 6936–6942. <https://doi.org/10.1145/3025453.3025804>
- [83] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: Automatic personalized health feedback from user behaviors and preferences using smartphones. In *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 707–718. <https://doi.org/10.1145/2750858.2805840>
- [84] Shriti Raj, Kelsey Toporski, Ashley Garrity, Joyce M. Lee, and Mark W. Newman. 2019. “My blood sugar is higher on the weekends”: Finding a role for context and context-awareness in the design of health self-management technology. In *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. <https://doi.org/10.1145/3290605.3300349>
- [85] Alvin Rajkumar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissam Hajaj, Michaela Hardt, Peter J. Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Yi Zhang, Gerardo Flores, Gavin E. Duggan, Jamie Irvine, Quoc Le, Kurt Litsch, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael D. Howell, Claire Cui, Greg S. Corrado, and Jeffrey Dean. 2018. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine* 1, 1: 18. <https://doi.org/10.1038/s41746-018-0029-1>
- [86] Meghan Reading Turchioe, Marissa Burgermaster, Elliot G. Mitchell, Pooja M. Desai, and Lena Mamykina. 2020. Adapting the stage-based model of personal informatics for low-resource communities in the context of type 2 diabetes. *Journal of Biomedical Informatics* 110, 103572. <https://doi.org/10.1016/j.jbi.2020.103572>
- [87] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender Systems: Introduction and Challenges. In *Recommender Systems Handbook*. Springer US, Boston, MA, 1–34. https://doi.org/10.1007/978-1-4899-7637-6_1
- [88] Darius A. Rohani, Andrea Quemada Lopategui, Nanna Tuxen, Maria Faurholt-Jepsen, Lars V. Kessing, and Jakob E. Bardram. 2020. MUBS: A Personalized Recommender System for Behavioral Activation in Mental Health. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3313831.3376879>
- [89] Hanna Schäfer. 2016. Personalized Support for Healthy Nutrition Decisions. In *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, 455–458. <https://doi.org/10.1145/2959100.2959105>
- [90] Hanna Schäfer, Santiago Hors-Fraile, Raghav Pavan Karumur, André Calero Valdez, Alan Said, Helma Torkamaan, Tom Ulmer, and Christoph Trattner. 2017. Towards Health (Aware) Recommender Systems. In *Proceedings of the 2017 International Conference on Digital Health - DH '17*, 157–161. <https://doi.org/10.1145/3079452.3079499>
- [91] Jessica Schroeder, Ravi Karkar, James Fogarty, Julie A. Kientz, Sean A. Munson, and Matthew Kay. 2018. A Patient-Centered Proposal for Bayesian Analysis of Self-Experiments for Health. *Journal of Healthcare Informatics Research*: 1–32. <https://doi.org/10.1007/s41666-018-0033-x>
- [92] Jessica Schroeder, Ravi Karkar, Natalia Murinova, James Fogarty, and Sean A. Munson. 2019. Examining Opportunities for Goal-Directed Self-Tracking to Support Chronic Condition Management. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4: 1–26. <https://doi.org/10.1145/3369809>
- [93] Jessica Schroeder, Chelsey Wilks, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M. Linehan. 2018. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2018)*: 1–15. <https://doi.org/10.1145/3173574.3173972>
- [94] Elizabeth Stowell, Mercedes C. Lyson, Herman Saksone, Renee C. Wurth, Holly Jimison, Misha Pavel, and Andrea G. Parker. 2018. Designing and Evaluating mHealth Interventions for Vulnerable Populations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–17. <https://doi.org/10.1145/3173574.3173589>
- [95] Kirsten Swearingen and R. Sinha. 2001. Beyond Algorithms: An HCI Perspective on Recommender Systems. *ACM SIGIR 2001 Workshop on Recommender Systems (2001)*: 1–11. <https://doi.org/10.1.1.23.9764>
- [96] Esteban G. Tabak and Giulio Trigila. 2018. Explanation of Variability and Removal of Confounding Factors from Data through Optimal Transport. *Communications on Pure and Applied Mathematics* 71, 1: 163–199. <https://doi.org/10.1002/cpa.21706>
- [97] United States Department of Agriculture (USDA). ChooseMyPlate. Retrieved September 16, 2020 from <https://www.choosemyplate.gov/>
- [98] Tiffany C. Veinot, Jessica S. Ancker, Heather Cole-Lewis, Elizabeth D. Mynatt, Andrea G. Parker, Katie A. Siek, and Lena Mamykina. 2019. Leveling Up. *Medical Care* 57: S108–S114. <https://doi.org/10.1097/MLR.0000000000001032>
- [99] Tiffany C Veinot, Hannah Mitchell, and Jessica S Ancker. 2018. Good intentions are not enough: how informatics interventions can worsen inequality. *Journal of the American Medical Informatics Association*. <https://doi.org/10.1093/jamia/ocy052>
- [100] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing theory-driven user-centric explainable AI. In *Conference on Human Factors in Computing Systems - Proceedings*, 1–15. <https://doi.org/10.1145/3290605.3300831>
- [101] M L Wheeler, A Daly, A Evert, and others. 2014. Choose Your Foods, Food Lists for Diabetics. *Chicago, IL: Academy of Nutrition and Dietetics/American Diabetes Association*.
- [102] Longqi Yang, Cheng-Kang Hsieh, Hongjian Yang, Nicola Dell, Serge Belongie, Curtis Cole, and Deborah Estrin. 2016. Yum-me: A Personalized Nutrient-based Meal Recommender System. *ACM Transactions on Information Systems* 36, 1: 7. <https://doi.org/10.1145/3072614>
- [103] David Zeevi, Tal Korem, Niv Zmora, David Israeli, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, Dar Lador, Tali Avnit-Sagi, Maya Lotan-Pompan, Jotham Suez, Jemal Ali Mahdi, Elad Matot, Gal Malka, Noa Kosower, Michal Rein, Gili Zilberman-Schapira, Lenka Dohnalová, Meirav Pevsner-Fischer, Rony Bikovsky, Zamir Halpern, Eran Elinav, and Eran Segal. 2015. Personalized Nutrition by Prediction of Glycemic Responses. *Cell* 163, 5: 1079–1095. <https://doi.org/10.1016/j.cell.2015.11.001>
- [104] Brian J Zikmund-Fisher, Aaron M Scherer, Holly O Witteman, Jacob B Solomon, Nicole L Exe, Beth A Tarini, and Angela Fagerlin. 2016. Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *Journal of the American Medical Informatics Association* 24, 3: ocw169. <https://doi.org/10.1093/jamia/ocw169>
- [105] Facts Up Front. Retrieved September 15, 2018 from <http://www.factsupfront.org/>

A APPENDICES

A CONTROLLED EXPERIMENT MATERIALS

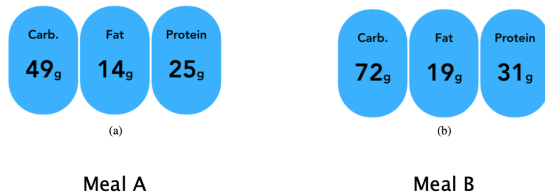
This section contains supplementary figures with example materials used in the controlled experiment.

2 Nutritional goal activity

A friend of yours with diabetes worked with their dietitian to set a nutritional goal. They asked you for help picking a meal that meets the goal.

Your Friend’s Goal: For high fat lunches, decrease your fat to be about $2\frac{1}{2}$ fat choices (12g). An example of 1 fat choice is $\frac{1}{4}$ medium avocado, 1 ounce of mozzarella, or $1\frac{1}{2}$ teaspoons of peanut butter.

Below is the nutrition information for two different meals. Please circle the nutrition information that you think is closer to your friend’s nutritional goal.



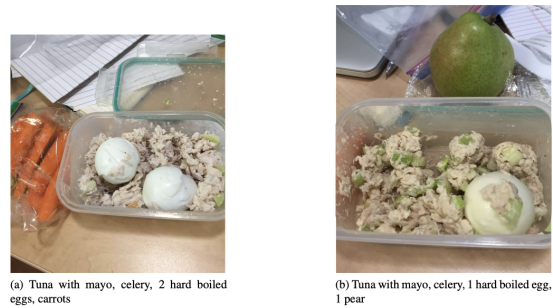
Supplementary Figure A. An example item from the goal comprehension task

7 Nutritional goal activity

A friend of yours with diabetes worked with their dietitian to set a nutritional goal. They asked you for help picking a meal that meets the goal.

Your Friend’s Goal: For high carb lunches, replace 1 carb choice with 1 heart healthy fat choice. For example, replace 1 apple with 1 ounce of feta or 1 ounce of mozzarella.

Below are images and descriptions for two different meals. Please circle the meal that you think is closer to your friend’s nutritional goal.



Supplementary Figure B. An example item from the goal/image matching task



Supplementary Figure C. The “virtual buffet” for breakfast meals.