

Good for the Many or Best for the Few? A Dilemma in the Design of Algorithmic Advice

GRAHAM DOVE, Tandon School of Engineering, New York University, USA

MARTINA BALESTRA, Tandon School of Engineering, New York University, USA

DEVIN MANN, Grossman School of Medicine, New York University, USA

ODED NOV, Tandon School of Engineering, New York University, USA

Applications in a range of domains, including route planning and well-being, offer advice based on the social information available in prior users' aggregated activity. When designing these applications, is it better to offer: a) advice that if strictly adhered to is more likely to result in an individual successfully achieving their goal, even if fewer users will choose to adopt it? or b) advice that is likely to be adopted by a larger number of users, but which is sub-optimal with regard to any particular individual achieving their goal? We identify this dilemma, characterized as *Goal-Directed* vs. *Adoption-Directed* advice, and investigate the design questions it raises through an online experiment undertaken in four advice domains (financial investment, making healthier lifestyle choices, route planning, training for a 5k run), with three user types, and across two levels of uncertainty. We report findings that suggest a preference for advice favoring individual goal attainment over higher user adoption rates, albeit with significant variation across advice domains; and discuss their design implications.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: advice applications; design dilemmas; empirical study

ACM Reference Format:

Graham Dove, Martina Balestra, Devin Mann, and Oded Nov. 2020. Good for the Many or Best for the Few? A Dilemma in the Design of Algorithmic Advice. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 168 (October 2020), 22 pages. <https://doi.org/10.1145/3415239>

1 INTRODUCTION

Applications that offer advice to users have become commonplace across domains as diverse as route planning (e.g. Google Maps or Citymapper), financial investment (e.g. PlanMode), and lifestyle and well-being (e.g. Noom). Increasingly, this advice is generated algorithmically, based on machine learning analysis of large amounts of historic data relating to aggregated prior use. The advice on offer thereby combines social cues from other users' activity with up-to-date data from the domain in question. Two important questions that may be considered when formulating this advice are: 1) is the user likely to adopt the advice, if it is offered? and 2) if adopted, is the advice likely to lead to the user achieving their goal?

Authors' addresses: Graham Dove, Tandon School of Engineering, New York University, New York, USA, gd64@nyu.edu; Martina Balestra, Tandon School of Engineering, New York University, New York, USA; Devin Mann, Grossman School of Medicine, New York University, New York, USA; Oded Nov, Tandon School of Engineering, New York University, New York, USA, onov@nyu.edu.

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 the author(s) 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.

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2020/10-ART168 \$15.00

<https://doi.org/10.1145/3415239>

In many cases, offering the advice most likely to result in users achieving their goal will also result in the highest rate of user adoption. However, this is not always the case. There are times when this advice may seem counter-intuitive; such as when traffic conditions result in a route much longer in distance being shorter in duration. In other situations, such as advising major changes in lifestyle associated with improving personal health, the ‘best’ advice might appear exceptionally challenging. For example, behavior in response to social distancing guidelines during the COVID-19 pandemic has highlighted the tension between the standalone value of public advice on one hand and its adoption by individuals who find it hard to follow on the other. Extant research shows that in cases like these people do not always act like models of rational decision making, and may not necessarily make decisions that maximize utility, or follow advice that is optimal to achieving their goal. Instead, their decisions are likely to be based on heuristics and judgments that follow predictable biases, e.g. [7, 28, 49, 50, 80, 85]. These circumstances pose a potentially serious design dilemma. Should the user be offered the advice which, if strictly adhered to, would be most likely to result in them achieving their objective, even if they are statistically less likely to adopt or adhere to it? Or alternatively, should the user be offered advice that they are more likely to choose to adopt and see through, and which may still be helpful, even though they are less likely to achieve their goal in full? We characterize this dilemma as being between: 1) *Goal-Directed advice* that is more likely to lead a user who fully adopts it to achieve their goal, even if overall adoption rates are likely to be lower, and 2) *Adoption-Directed advice* which is likely to have a higher adoption rate, but which has a lower probability of resulting in a user fully achieving their individual goal.

This dilemma is brought even more clearly into focus when machine learning and data analysis techniques used to predict likely outcomes might also be used to predict user adoption rates. What might this mean for designers of applications aiming to algorithmically generate valuable advice in complex situations? Which design choice would be preferable from the perspective of a user? Which type of advice might be considered more ethical? This paper reports on a study in which participants in an online experiment were asked to indicate a preference between examples of these different advice types in one of four different scenarios, each presenting the potential dilemma in a different domain context with particular characteristics that influence how the dilemma might present. We varied the role participants were asked to adopt when indicating their preference, and controlled for several other potentially confounding factors. Our analysis found that participants consistently favored presenting Goal-Directed advice over Adoption-Directed advice, albeit the degree to which this preference was shown differed significantly across the scenarios that were presented. We discuss our findings with reference to prior research in decision-making under uncertainty, recommender systems, and behavioral economics, and outline implications for designing systems that offer algorithmically generated advice.

1.1 The Advice Design Dilemma

The research reported in this paper addresses a dilemma in presenting users with algorithmic advice that aims to support their decision making under uncertainty, particularly in challenging contexts. This dilemma can appear when advice, which may be optimal to an individual achieving their goal, appears counter-intuitive. Or when that advice seems challenging to such a degree that many users may either discount or ignore it. In these circumstances, is this advice still optimal? and should an application offer this advice? Or is it preferable for an application to offer advice that a higher percentage of users are more likely to adopt, but which is less likely to lead these users towards fully achieving their goal?

For the purposes of this study we define these two alternative advice types as follows:

- *Goal-Directed advice*: advice that is more likely to result in users who adhere to it achieving their goal, but which users are less likely to adopt and adhere to.
- *Adoption-Directed advice*: advice that is more likely to be adopted and adhered to by a greater number of users, but which is less likely to result in a user fully achieving their goal.

It should be noted that in the situations discussed in this paper, we make explicit the assumption that selecting either of the advice options is preferable to acting independently, and would therefore result in a better outcome for the user than were they to follow neither.

1.2 Contributions

This research contributes to the CSCW community's understanding of how to design applications that offer algorithmically generated advice based on social information aggregated from the prior behavior of a large number of users, and has both theoretical and practical implications. From a theoretical perspective, we 1) extend prior research at the intersection of CSCW and recommender systems that offer socially-informed advice based on aggregated data from prior use; 2) introduce and characterize the design dilemma posed when having to select between Goal-Directed and Adoption-Directed advice; and 3) quantify the trade-offs it represents. From a practical perspective we offer initial guidance on how to approach this dilemma when designing applications that offer algorithmically generated advice.

2 RELATED WORK

This research was inspired by the observation that users may discount or ignore algorithmic advice intended to support the choices they make towards achieving particular goals, specifically when this advice may appear counter-intuitive or challenging to adhere to. In order to situate our work we provide an overview of related prior research in 1) decision making under uncertainty; and 2) online advice and recommendation.

2.1 Decision-making under uncertainty

A primary motivation for offering algorithmic advice is to support users' decision-making. Under uncertainty, decision-makers tend to be receptive to advice [14, 90], from experts [25, 48, 64, 86] and from peers [20, 32, 81]. This tendency has informed the design of interfaces for recommender systems. For example, in financial planning [2, 87] advice from software agents is increasingly replacing advice from humans [3, 46], evidenced in the growth of financial 'Robo-Advisors', such as Betterment¹, which provide online advice with minimal human intervention.

Even where valuable advice may be available, decision-makers still tend to rely on simple heuristic principles to reduce the complexity of tasks such as assessing probabilities and predicting values [7, 50]. While often useful, heuristics can also lead to systematic errors or biases [85]. Their impact has been extensively studied in the context of personal finance [10, 81]. Prior research shows the effectiveness of changing default participation to opt-in, but also the negative impact a reliance on simplistic heuristics has on asset allocation [9]. Decision-makers may excessively discount advice even when shown that the advice is good [36, 60], because they have access to their own but not others' justifications [90], or because of a bias toward egocentric or self-related information [26, 30, 54, 55]. As a result, decision-makers tend to exaggerate their own abilities [55], not take others' skills sufficiently into account, and display overconfidence and unrealistic self-assessments [13]. However, task complexity and quality of explanation may provide a counter-balance that reduces advice discounting [70], and decision-making can be affected by context driven challenges to self-control, which may represent inconsistent long-term and short-term preferences [78, 80].

¹www.betterment.com

2.2 Online Advice and Recommender Systems

Insights such as these, from behavioral economics and social psychology, have influenced HCI, CSCW and information systems research (e.g. [21, 38, 58]) in areas such as recommender systems [1, 79]. Recommender systems present advice to users in the form of suggestions [67, 88] often based on collaborative filtering and a social context [44, 84]. Recommender systems have been used to provide advice in a wide range of domains and use cases [46], including advice on energy saving [5], nutrition [91], finance [39], and remedial behavior for computer programmers [42]. Users' interactions with such systems are susceptible to similar confirmation biases as other instances of decision-making and advice selection [47], which has led researchers to investigate diverse recommendation and dissenting information strategies [19, 68]. In addition, prior research at the intersection of CSCW and recommendation systems found that negative social effects of activity transparency can result in users' increased adoption of mediocre advice [67]. Another significant challenge to overcome, identified in this line of research, is the tendency for users to be put off by recommendations they consider demanding or challenging [77]. Moreover, the recommendations users find most useful may not always be the ones that by other measures are considered most accurate [62]. An example is when optimal advice appears counter-intuitive, but following sub-optimal advice will still result in increased profit or satisfaction [59]. We extend this prior research by investigating participants' preferences between Goal-Directed and Adoption-Directed advice, under uncertainty and in counter-intuitive or challenging scenarios.

3 EXPERIMENTAL STUDY

To investigate the advice design dilemma, we probe participants' design preferences for which advice type to present in a future app. To study this, we conducted a between-subjects experiment online using Amazon Mechanical Turk (MTurk) as a recruitment platform. While acknowledging that the dilemma may appear in situations where there are more complex, ongoing interactions with applications that reoccur over time, and involve both short-term and evolving longer-term goals, e.g. when using diet and exercise apps, we chose to echo a long tradition in economics, where simple, discrete-choice experiments are used to isolate a particular issue of concern and elicit preferences, e.g. [6, 17, 31, 45]. This degree of simplification allows us to gain purchase on the novel conceptual ground that the dilemma reflects, by allowing us to isolate the particular cases where advice might appear counter-intuitive or challenging, and to investigate the impact of factors, such as task domain, on these instances.

3.1 Research Questions

We ask the following research questions:

- *RQ1: Are participants' preferences between offering Goal-Directed advice and Adoption-Directed advice sensitive to different domain scenarios, or do they transcend specific settings?*
- *RQ2: Are participants' preferences between offering Goal-Directed advice and Adoption-Directed advice sensitive to the different perspectives of 'advice giver' and 'advice receiver'?*

3.2 Study Design

We recruited a total of 1,589 US MTurk participants over the age of 18. Participation was limited to people with a record of at least 100 tasks at an approval rate of 95% or higher. Of these, 750 self-identified as women and 829 as men. Their self-reported ages ranged between 18 and 88, with a mean age of 36 (median = 33). Participants were paid a flat rate of \$0.65 for completing the survey based on an estimated time for completion of 5 minutes (median time spent on the study by participants was 5 minutes 30 seconds), and could take part in the study only once. Each participant

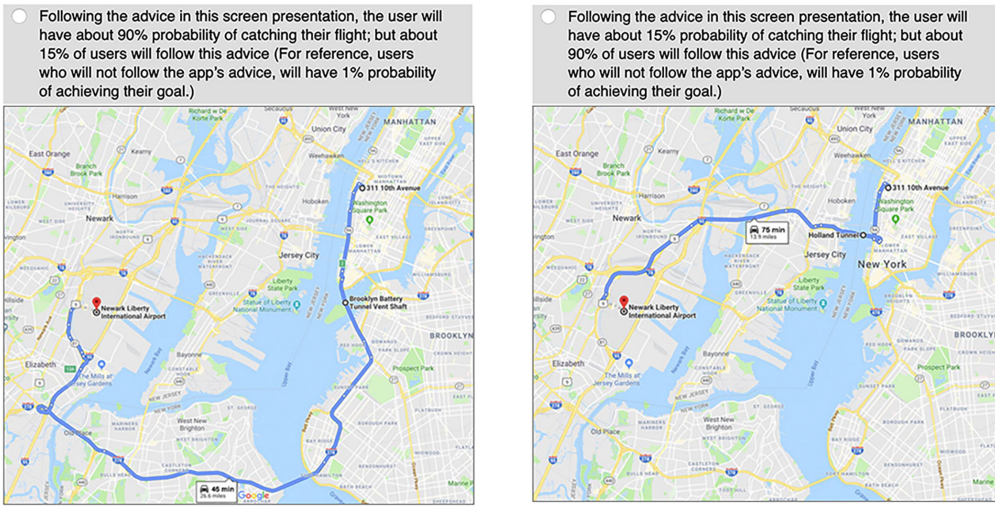
was presented with a single selection task, asking for their preference between Goal-Directed and Adoption-Directed advice. Based on a 4X3 factorial design, participants were randomized into one of four scenarios, each set in a different domain, and three roles through which to consider their selection. We also controlled for other potentially confounding factors, such as the size of the gap between adoption and goal-effectiveness probabilities, and order of presentation. In addition to making their advice type selection, participants were required to answer an attention test question. Finally, participants were asked to briefly explain the reasons behind their selection. In the following sections we report the details of experimental procedure.

3.3 Domain Scenarios

To address RQ1, we selected four different domain scenarios representing different contexts where users might seek advice. We designed each scenario to present a situation in which the advice most optimal for individual goal attainment is likely to appear counter-intuitive or seem challenging to adhere to. In these scenarios, we selected the details of both advice types based on realistic examples of advice offered in similar situations (sources referenced in the descriptions below). These scenarios were:

- *Route planning*: In this scenario (Figure 1a), participants were asked to consider the case of a driver aiming to catch a flight and using an app to direct them to an airport in an unfamiliar location. It was based on examples of similar route planning (e.g. using Google maps) under different traffic conditions.
- *Investment planning*: This scenario described a novice investor planning for retirement and aiming to recover from the previous year's poorly performing stock market. Two investment plans were shown. It was based on previous research into investment planning for retirement [37, 38].
- *Training for a 5k run*: The third scenario presented a first time runner using a fitness-training app to guide their preparation for a 5K race in which they aim to finish in less than 26 minutes. It was based on training advice from [35, 56].
- *Making healthier lifestyle choices*: The fourth scenario described a user planning healthier lifestyle choices following a medical checkup. It was based on exercise and diet guidelines from [69].

We selected four different domains in order to test whether participants' choices are sensitive to domain. The domains selected represent typical areas where algorithmic advice might be offered, but each has particular characteristics that make them interestingly different. For example, we selected 'route planning' because the goal is immediate, whereas 'training for a 5k run' offers a more long-term but well defined goal. In the 'making healthier lifestyle choices' scenario the goal is both longer-term and less concrete, while we selected an 'investment planning' scenario where the goal is in the longer-term future but decisions are made in response to historic activity. In each of these scenarios, participants were presented with a choice of two screen designs (e.g. Figure 1). One screen shows Goal-Directed advice that is either counter-intuitive (e.g. in the route planning scenario the driver is required to double back and drive a significantly longer distance) or challenging to adhere to (e.g. in the making healthy lifestyle choices scenario the plan set strict rules for diet and exercise). Participants were told that fewer people were likely to follow the advice presented on this screen, but that if they did manage to adhere to it, they would have a greater chance of achieving their goal. On the second screen design we presented Adoption-Directed advice that might appear more immediately intuitive (e.g. in the route planning scenario the route was shorter in distance and appeared more direct) or easier to adhere to (e.g. in the making healthier lifestyle choices scenario the plan set looser, less stringent, targets). Participants were told that more



(a) ‘Goal-Directed’ advice: Participants are shown a route that is longer in distance but shorter in duration. They are instructed that users will have a 90% probability of catching their flight, but that only 15% of users are likely to adopt it.

(b) ‘Adoption-Directed’ advice: Participants are shown a route that is shorter in distance but longer in duration. They are instructed that users will have a 15% probability of catching their flight, but that 90% are likely to adopt it.

Fig. 1. Screen designs from the ‘Route planning’ scenario, where the goal-adoption gap is 90-15 and the ‘goal’ term is given first in the description text. We show, (a) Goal-Directed advice and (b) Adoption-Directed advice. The text above each image explains the advice-type choices to participants.

people were likely to follow this advice, but that they would have a reduced chance of completely achieving their goal.

3.4 Participant Roles

To address RQ2, and investigate whether preferences would manifest differently depending on whether participants were offering or receiving advice, we randomly assigned each participant to one of three roles, *developer*, *user*, and an *ethical choice* role. By assigning participants the role of developer we placed them in a scenario where information about likely rates of goal success and adoption would be available, therefore highlighting the dilemma as it may manifest in realistic settings. These participants were asked, “If you were the app developer, what would you show to the user in this situation?”. In assigning participants the role of user they were placed in a setting akin to a user study where they were presented with goal success and adoption information to elicit a hypothetical choice. These participants were asked, “If you were the app user, what would you want to be shown in this situation?”. By assigning participants the ethical choice role we were presenting a scenario where they would need to know about the potential for individual goal success and likely adoption rates in order to weigh up the ethics of their advice type preference. These participants were asked, “Which version of the screen below is more ethical advice to give to the user?”. We did not provide a definition of what ‘ethical’ should be, rather preferring to leave this to participants’ own conceptualization of the term. We did this in order to allow participants to focus on responding to the dilemma posed, rather than potentially taking issue with an externally imposed understanding of what is ethical. Our advice choice rationale question, discussed below,

allowed us the opportunity to unpick possible issues that might arise from participants' different conceptualizations.

3.5 Additional Study Design Considerations

3.5.1 Goal-adoption gap. We also wanted to account for the possibility that participants' preferences would be influenced by the apparent size of the gap between the probability that users would achieve their goal and the probability that users would follow the advice. To control for this, we decided to create two variations and randomly assign participants to one of these. In the first, participants selected between a Goal-Directed advice option with a 90% goal attainment probability but 15% adoption rate, and an Adoption-Directed option with 15% goal attainment probability but 90% adoption rate. In the second variation, the Goal-Directed advice option presented 60% goal attainment probability but 45% adoption rate, and the Adoption-Directed option presented 45% goal attainment probability but 60% adoption rate. In all cases participants were told that choosing to follow *neither* advice would result in the user having only a 1% probability of achieving the goal. This was to highlight to participants that both advice options would be valuable to the user.

3.5.2 Ordering effects of interface elements. To control for possible ordering effects, we randomly varied the presentation of advice type options on the page, such that half the time Goal-Directed advice was on the left-hand side and Adoption-Directed advice was on the right, and the other half this was reversed. We also randomly varied the order of the terms referencing Goal-Directed and Adoption-directed advice in the description text above the screen designs. Half the time, the Goal-Directed probability was presented first, e.g. in the route planning scenario with a 90-15 goal-adoption gap the descriptions would read: "Following the advice in this screen presentation, the user will have about 90% probability of catching their flight; but about 15% of users will follow this advice". For an example see Figure 1. In the remainder, the Adoption-Directed probability was presented first, e.g. in the investment scenario with a 60-45 goal-adoption gap the descriptions would read: "About 60% of users will follow this advice; but following the advice in this screen presentation, the user will have about 45% probability of reaching their retirement goal".

3.5.3 Attention test. To filter out participants who simply rushed through the study without taking time to consider their responses, we also included a simple attention test question on the main study page based on the content of the particular scenario being shown. This type of test is commonly added to experimental tasks undertaken on MTurk.

3.5.4 Independence of goal attainment and adoption rate probabilities. Because we were asking participants to make a design decision regarding which advice type to present to users, rather than which advice they would follow themselves, we wanted to be confident that participants understood the trade-off the dilemma poses. One important aspect of this was understanding that the predicted goal attainment and likely adoption rate probabilities were independent of each other, rather than on an interconnected 'sliding-scale', such that an increase in one would automatically lead to an equivalent reduction in the other. To test for this, we added a multiple choice question set in a different domain scenario than the participant had been presented in the main study question. Here we presented participants with a single screen design, saying that it would offer users a 50% probability of achieving their goal. We then asked them to select from 4 choices the proportion of users who would likely follow this advice (the correct response being: 'The scenario does not provide enough information to answer', see Figure 2).

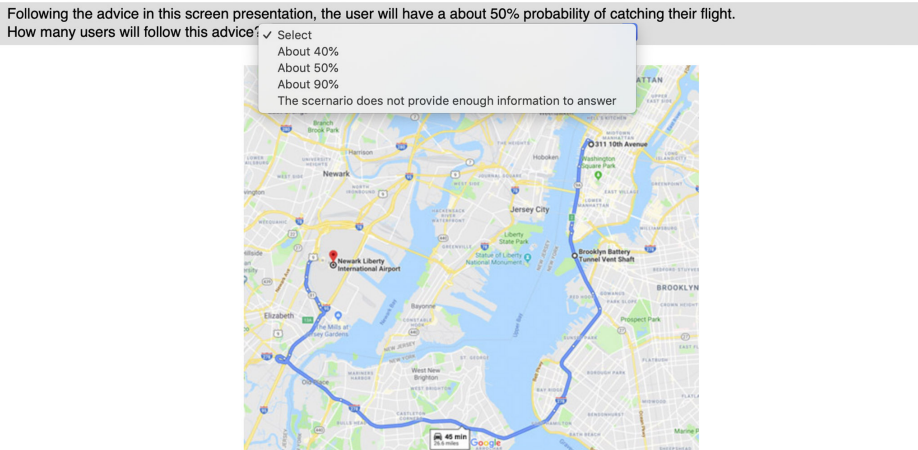


Fig. 2. Screenshot of the second scenario test of participants' understanding of independence between the goal attainment and adoption rate probabilities in the 'Route planning' scenario.

3.6 Advice choice rationale

To better understand participants choice between advice types, and to probe the reasons behind these preferences, we also asked them to briefly explain their choice. First we reminded participants of the selection they had made in the main scenario, and then we asked them for a brief free-text response explaining why they selected their chosen advice type. We also probed participants about their responses to the second scenario in a similar fashion, using these responses to sanity check our decision to test for an understanding of goal attainment and adoption rate independence.

4 ANALYSIS

In order to have more confidence in our findings, we adopted a conservative approach to data inclusion. First we discounted 3 submissions from participants who did not complete the task or who spent less than one minute on the task, as we considered these unreliable. We then discounted data from 434 participants who responded incorrectly to the attention check question, the answer to which could be readily found in the text describing the scenario. Such attention tests are commonly used to address potential concerns over data validity in studies where participants are recruited through MTurk [43, 57], and this is not an atypical number, as up to 42% of participants in MTurk studies have been found to be inattentive [33]. We then also removed data from 444 participants who responded incorrectly when tested about the independence of advice type probabilities. Data from these participants was discounted because we felt that a failure to understand the independence of the probabilities shown for Goal-Directed and Adoption-Directed advice indicated that participants may not clearly understand the choices available when the dilemma presented. This is a relatively large number of participants to exclude, and we speak further to this decision in the Discussion section. After the removal of data from 772 of 1,589 candidates, we were left with a final data set of 817 participants.

For our quantitative analysis, we first separated participants' response data according to their role assignment, i.e. user, developer, or ethical choice; and the scenario they were assigned, i.e. 'route planning', 'investment planning', 'training for a 5k run', or 'making healthier lifestyle choices'. The number of participants in each category is recorded in Table 1.

	Scenario			
Role	Healthy Living	Investment Planning	Race Training	Route Planning
developer	68	58	76	88
ethical	74	73	55	81
user	54	55	53	82

Table 1. Number of participants in each experimental category.

We then compared the rate at which participants in each sub-group preferred the Goal-Directed advice to a 0.5 ‘indifference’ threshold, using one-sample z-tests. A 0.5 reference threshold was used for comparison because we would expect participants to be *indifferent* if 50% of the sample selected the goal-directed advice and 50% selected the adoption-directed advice. This initial analysis provides us with a high level view of the preferences of each of these sub-groups.

To extend our analysis, we used a logistic regression to investigate how these variables might influence an individual’s probability of selecting the Goal-Directed advice. To do this we created three models. In Model 1 we included the participant’s assigned role and scenario as main features in the regression. For Model 2 we also included terms for covariates: the goal-adoption gap, the side of the page on which the advice types were presented, the order that advice types were listed in the description text, participant gender, the log of participant age, and the log of time taken to complete the study. Finally, in Model 3, we included an interaction term between participant role and scenario. We use logistic regression because our dependent variable and the majority of our independent variables were categorical.

We also analyzed these participants’ responses to the post-study question probing their advice choice rationale. We received responses from 814 of the 817 participants included in our quantitative analysis. These responses were brief, typically somewhere between a few words and a short paragraph in length, with a mean response length of 24 words. There was no need to sanitize these responses, most likely due to conservative approach we took to data inclusion (described above). We performed a simple thematic analysis [16], which involved an initial close reading and clustering led by the first author, followed by refinement and agreement seeking with one other researcher. For a first pass, the responses from all participants were considered together, resulting in a high-level set of main themes. Following this, we divided the responses into two groups so that those who selected a preference for Goal-Directed advice and those who selected Adoption-Directed advice were considered independently. This provided us with a more fine-grained understanding of the differences between the two groups and similarities within them. Having identified patterns across the data, and extracted key ideas, we sought agreement across interpretations, through discussion and where necessary testing alternative framings.

5 FINDINGS

Of the 817 participants in our final data set, 658 (80.54%) selected a preference for displaying Goal-Directed advice over Adoption-Directed advice ($z=17.46$, $p<0.05$, 95% C.I.[0.78,0.83]). This preference was consistent across all scenarios and all roles, and in all cases this preference was statistically significant using a one-sample z-test of proportions relative to the 50% indifference threshold. With regard to the different domain scenarios, we found that 74.49% of participants shown the ‘making healthier lifestyle choices’ scenario ($z=8.31$, $p<0.05$, 95% C.I.[0.70,0.79]), 80.64% of participants shown the ‘investment planning’ scenario ($z=10.72$, $p<0.05$, 95% C.I. [0.77,0.84]), 74.45% of participants shown the ‘training for a 5k run’ scenario ($z=8.24$, $p<0.05$, 95% C.I. [0.70,0.79]), and 89.64% of participants shown the ‘route planning’ scenario selected a preference for Goal-Directed

advice ($z=15.43$, $p<0.05$, 95% C.I. [0.87,0.92]). These rates are visualized in Figure 3a. Similarly, 78.28% of participants assigned the role of developer ($z=11.69$, $p<0.05$, 95% C.I. [0.75,0.82]), 83.04% of those assigned the ethical choice role ($z=13.97$, $p<0.05$, 95% C.I. [0.80,0.86]), and 80.32% of those assigned the role of user ($z=11.87$, $p<0.05$, 95% C.I. [0.77, 0.84]) selected a preference for Goal-Directed advice (Figure 3b).

Extending beyond these high level findings, the results of the logistic regression (Table 2) indicates that the scenario participants were shown significantly influences the probability that they select a preference for Goal-Directed advice. In particular, Models 1 and 2 show that participants shown the ‘route planning’ scenario were significantly more likely to select this preference relative to the other three scenarios. By converting the coefficients on scenario from Model 2 to probabilities (holding age and time on task at their medians - 35 years and 312 seconds, respectively), we see that the predicted probability of an individual selecting Goal-Directed advice is 94.23% when shown the ‘route planning’ scenario. This is significantly higher relative to the other scenarios: 84.39% in the ‘making healthier lifestyle choices’ scenario ($\beta = -1.11$, S.E. = 0.27, $z=-4.14$, $p<0.000$, 95% C.I. [-1.63,-0.58]), 88.55% in the ‘investment planning’ scenario ($\beta=-0.75$, S.E. = 0.28, $z = -2.67$, $p=0.008$, 95% C.I. [-1.30,-0.20]), and 84.79% in the ‘training for a 5k run’ scenario ($\beta=-1.08$, S.E. = 0.27, $z = -3.99$, $p=0.0001$, 95% C.I. [-1.60,-0.55]). These comparisons are significant even when correcting for multiple comparisons using a Bonferroni correction with a threshold of $p=0.017$. On the other hand, Model 3 shows that participants’ did not differ in their preference for Goal-Directed advice across roles.

Findings from our analysis of participants’ responses to the post-study choice rationale question, explaining *why* they selected a preference for presenting users with Goal-Directed advice or Adoption-Directed advice, offer further insight into these quantitative results. A number of examples from this analysis are included in the discussion that follows, here we briefly list the themes that emerged.

- *Relationships between individual users and the crowd:* Some participants selecting a preference for Goal-Directed advice explained their choice in terms of *supporting an individual user in*

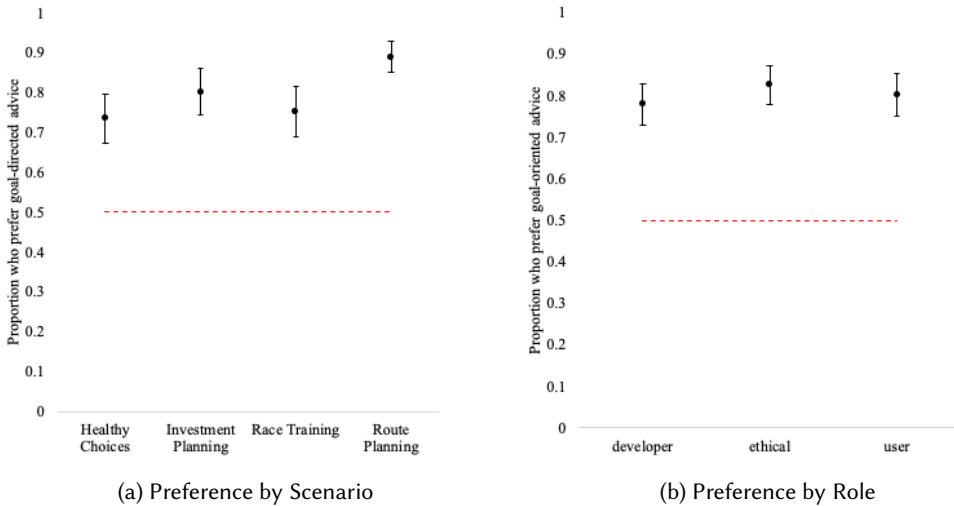


Fig. 3. Proportion of people preferring Goal-Directed advice according to: (a) the scenario presented; and (b) the role they were assigned, with 95% confidence intervals. Red lines represent the 0.5 ‘indifference’ threshold

Table 2. Results of logistic regression testing effect of role and domain on probability of selecting Goal-Directed advice. Model 1 is the model of main effects; Model 2 includes covariates; Model 3 includes an interaction effect between role and domain.

		Model 1	Model 2	Model 3
Intercept		2.07(0.26)***	3.26(1.44)*	3.99(1.51)**
Role	<i>developer</i>	-0.08(0.22)	-0.07(0.22)	-0.54(0.54)
	<i>ethical</i>	0.22(0.22)	0.24(0.23)	-0.46(0.56)
Domain	<i>health</i>	-1.11(0.26)***	-1.11(0.27)***	-1.45(0.53)**
	<i>investment</i>	-0.75(0.28)**	-0.75(0.28)**	-1.45(0.53)**
	<i>race</i>	-1.08(0.27)***	-1.08(0.27)***	-1.51(0.53)**
Gap size	<i>90-15 split</i>		0.11 (0.18)	0.10(0.18)
Text order	<i>goal first, adoption second</i>		0.09(0.18)	0.07(0.18)
Screen side	<i>goal left, adoption right</i>		-0.09(0.18)	-0.10(0.18)
Gender	<i>male</i>		-0.21(0.18)	-0.19(0.19)
	<i>other</i>		-0.04(1.15)	-0.05(1.15)
Age			-0.41(0.31)	-0.45(0.31)
Time on task			0.07(0.19)	0.03(0.19)
Role*Domain	<i>developer*health</i>			0.42(0.68)
	<i>developer*invest.</i>			1.16(0.72)
	<i>developer*race</i>			0.31(0.67)
	<i>ethical*health</i>			0.60(0.70)
	<i>ethical*invest.</i>			0.94(0.71)
	<i>ethical*race</i>			1.11(0.73)
AIC		793.08	801.59	808.00
Pseudo R ²		0.033	0.037	0.044

(Signif. codes: < 0.001"***"; 0.001 – 0.01"***"; 0.01 – 0.005"***"; 0.05 – 0.1"***")

successfully achieving their goal. Others stressed that it was *the responsibility of individuals whether they choose to follow advice or not, or wanted to reward those with the capacity to adhere to advice*, or adopted an *egocentric standpoint* stressing their own qualities in comparison to others. For participants who selected Adoption-Directed advice, *concern for the wider benefits of a greater number of people* or the *pragmatic partial achievement of a goal by larger numbers of users* were important explanations.

- *Personal insight and domain experience*: This was a strong theme among participants selecting Adoption-Directed advice, who made *decisions based on heuristics* such as ‘diversify investment’. Those selecting Goal-Directed advice often made *connections to their own lifestyles or practices*.
- *Information detail and complexity*: For those selecting Goal-Directed advice, the *discipline imposed by a more detailed plan* would make it easier to follow through to successful goal achievement. Participants selecting a preference for Adoption-Directed advice suggested the less detailed and demanding option would be *a more realistic plan*.
- *Success of the app or company*: Typically found among participants in the developer role, but across all scenarios and both advice types, this theme stressed success of the app, or the company behind it. For participants selecting Adoption-Directed advice the focus was on attracting a high number of downloads and *large initial user base*, so that the app would be a

viable business proposition. For participants selecting Goal-Directed advice the focus was more on *user retention and satisfaction*, on minimizing bad reviews and increasing good ones.

6 DISCUSSION

Our findings show participants indicating a clear preference for offering users Goal-Directed over Adoption-Directed advice, in the tasks we presented. This was consistent, albeit with varying rates of preference, across each of the four scenarios we tested in order to address RQ1, and was also invariant to the role participants were assigned in order to address RQ2. This suggests that most participants focused on those users able to adopt and then adhere to the advice being offered, rather than the wider population. Such a finding is consistent with prior findings about moral dilemmas, (e.g. [15, 82]), which suggest that people generally avoid decisions that are sub-optimal at the level of individual utility, even if they increase aggregated social utility. Our findings about participants' choices are echoed in our analysis of participants' responses to the advice choice rationale question, where one of the themes to emerge under the theme *'Relationships between individual users and the crowd'* was *'supporting an individual user in successfully achieving their goal'*, while another emphasized that it is *'the responsibility of individuals whether they choose to follow advice or not'*, e.g.:

"The point of investment advice is to achieve the goal - it is up to the individual to decide if they will follow the advice" (*Investment planning, user, 90-15, Goal-Directed advice*)

"The app can't control who chooses to follow their advice, but they can control the quality of that advice, so I felt it was more ethical to provide investment advice that had a higher likelihood of helping a committed client reach his/her investment goal." (*Investment planning, ethical, 90-15, Goal-Directed advice*)

Such views were not universal though. A significant number of participants selected Adoption-Directed advice, and the rationale question responses indicate that *'concern for the wider benefits of a greater number of people'*, which might accrue through the *'pragmatic partial achievement of a goal by a large number of people'*, were key motivators. This was fairly common among participants presented with the 'investment planning' and 'making healthy lifestyle choices' scenarios, but rarer among those shown the 'training for a 5K run' scenario; and there were no examples at all amongst those in the 'route planning' scenario. Example responses include:

"Getting more people to consider their investment goal is still very beneficial, even if they fall short of their goal. This scenario meant that more people would be thinking about and taking action with their investments." (*Investment planning, developer, 60-45, Adoption-Directed advice*)

"I think it would help more people make some steps toward a healthier lifestyle which is better to me than fewer people achieving completely healthy lifestyles." (*Making healthy lifestyle choices, developer, 60-45, Adoption-Directed advice*)

These different preference explanations are one indicator of the influence scenario may play in the selection of advice types; suggesting greater nuance underneath the across the board preference we found in response to RQ2. Another is our quantitative finding that participants shown the 'route planning' scenario were significantly more likely to select a preference for Goal-Directed over Adoption-Directed advice. Each of these points to the importance of task complexity and task familiarity in the decisions participants made.

6.1 Task complexity

The ‘route planning’ scenario is arguably simplest, and most likely to manifest in the wild as a one-shot binary-selection. In this scenario, success only benefits the individual user, and there can be no partial achievement of the goal. The user either arrives in time to catch their flight, or they do not. Because of this, research showing the importance of loss aversion in decision-making e.g. [7, 50] might offer an explanation for our finding that participants shown this scenario were significantly more likely to prefer Goal-Directed over Adoption-Directed advice, when compared with participants shown any one of the other three scenarios. In contrast, the ‘making healthier lifestyle choices’ scenario is likely the most complex when translated into real world experiences. Yet here too task complexity played a role. For participants preferring to offer Goal-Directed advice, the ‘*discipline imposed by a more detailed plan*’, and the specific targets set would make it easier to follow through to successful goal achievement, e.g.:

“The instructions were much more rigid and difficult to achieve. For example, where one plan said to simply reduce alcohol consumption, the other listed an exact number to stay below. I felt that people that followed the more detailed plan received a better outcome than following a very basic and vague outline of a plan.” (*Making healthy lifestyle choices, user, 90-15, Goal-Directed advice*)

However, participants shown this scenario but preferring Adoption-Directed advice suggested the less detailed and less demanding option would be simpler to understand and ‘*a more realistic plan*’, that would also be easier to adhere to. This supports research suggesting users are put-off following advice from online recommender systems if it is considered demanding or challenging, e.g. [77].

“Based on what I’ve seen, most people would take one look at the other plan and run away from it. There’s too much to it and people don’t want to be bothered with programs like that. I think more people would be willing to take the common sense approach, which is what the plan I chose was.” (*Making healthy lifestyle choices, developer, 60-45, Adoption-Directed advice*)

Further research is needed to better understand the impact of task complexity on participants’ preferences. In this study, the choice between Goal-Directed and Adoption-Directed advice was presented as a binary selection between static screen designs. While this was intentional, and allowed us to isolate particular issues of concern, additional research is needed to fully investigate the nuances within more complex evolving interactions. Future studies should consider a longitudinal approach, that allows a deeper dive to study of how different options and trade-offs might be captured and expressed dynamically for users to explore and select between.

6.2 Task familiarity

In addition to being simplest, the ‘route planning’ scenario may also be the most familiar to participants. The use of journey planners is widespread, and research indicates that in circumstances where travel time is the key factor users are likely to follow their advice [76]. For the large number of participants who selected a preference for Goal-Directed advice when presented with this scenario, it may be that trust in this technology, and familiarity with the visualizations associated with it, reduces the apparent counter-intuitiveness of the advice offered. Further evidence of the impact that task or scenario familiarity might have on participants’ advice type preferences comes via references to ‘*personal insight and domain experience*’ in their explanations of advice type selections. This was evident among participants that selected Adoption-Directed advice who often made ‘*decisions based on heuristics*’, for example:

“Because historically, placing all your eggs in one basket is bad. Better to spread the risk and be more diversified. A fund that does well in the previous year will not necessarily do well in the future.” (*Investment planning, ethical, 90-15, Adoption-Directed advice*)

While those who selected Goal-Directed advice often made ‘connections to their own lifestyles or practices’, such as:

“I don’t mind taking back roads or scenic detours as long as I get to my destination on time. I know big city traffic can be horrible especially during rush hours and on popular roads. Sometimes taking a round about route gets you to your destination faster.” (*Route planning, user, 90-15, Goal-Directed advice*)

We suspect that participants make these personal references for reasons closely related to those underlying research that suggests decision-makers privilege self-related information and adopt an egocentric standpoint, e.g. [26, 30, 54, 55], and to research suggesting decision-makers may discount advice because they have access to their own justifications, but not to those of others [90]. Because of this, future studies might investigate how advice types can be personalized.

6.3 Cultural differences

Participants’ average preference for Goal-Directed over Adoption-Directed advice may also reflect dominant attitudes towards individual achievement versus collective benefit in the society our sample is drawn from. We recruited MTurk workers with an account in the U.S, a society where individualism is highly valued [61, 83]. It remains an open question whether our findings would translate to participants recruited from societies considered to place greater value on shared responsibility [40] or collectivism [61, 83]. In addition to potential biases associated with national cultures, there may also be a bias in the culture of participation in MTurk. Paid crowd workers may make different choices from participants recruited in other ways. Further research should investigate this, as potentially analogous behavior has been seen in comparisons between tasks undertaken by crowd workers and citizen scientists [22].

This potential bias could explain the many examples in which participants who preferred Goal-Directed advice adopted an ‘egocentric standpoint’ in their explanation that was explicitly made in opposition to others, or which stressed the individual qualities of the participant themselves, e.g.:

“I’m after the results. I don’t care whether others are willing to put in the effort - I am.” (*Making healthy lifestyle choices, ethical, 60-45, Goal-Directed advice*)

“I don’t need to feel validated by others’ choices. If I know I have a higher chance of reaching my goal I’m going to take it.” (*Route planning, user, 90-15*)

“I am a goal oriented person, once I start a goal I almost ALWAYS reach it. So for me, that scenario was more appealing.” (*Training for a 5k run, developer, 90-15, Goal-Directed advice*)

However, statements such as these may also reflect previous research which suggests decision-makers make overconfident self-assessments [13] and overestimate their own abilities [55]. It could be that participants who are able to simply ‘make a change’, e.g. when dieting, show a preference for Goal-Directed advice and simply assume that others are similar. Further research is needed to unpack the more nuanced impact of culture on peoples’ advice type preferences, as design decisions based on social cue data gathered in a single country, or from within a single online community, may not translate into other contexts.

6.4 Design values and ethical considerations

In making decisions about what type of advice is best to offer users, designers are asked to make value judgments in complex and ethically ambiguous situations. These judgments, and the values they operationalize, may be subject to different interpretations when employed in the service of an altruistic cause than when used to drive commerce [24]. Yet, despite the influence designers have on peoples' lives [18, 34, 75], prior discussion of the role of ethics in the design of technologies that offer advice or try to persuade users has tended to focus on the role they might play in the physical [72], material [37], and psychological welfare [52] of individuals, considering societal good as reflecting value judgments on what might be considered 'good' for the individual user, e.g. [11], rather than possible wider impacts. Excepting 'route planning', in each of the scenarios we presented there are potential benefits for users who partially achieve their goal. These benefits would be additional to the benefit gained by those who ultimately achieve their goal in full. The ambiguity surrounding such judgments is also reflected in themes from our analysis of the choice rationale responses. This highlights important ethical questions for designers who may have to choose between showing people what they apparently want as individuals, versus showing them what may have a larger positive impact on the greatest number of people. This contrast, between the individual and the group, adds an additional layer of complexity to decisions about 'good for whom', in a similar way to how feminist HCI [8] and HCI for sustainability [12, 29] have called on designers to reconsider what it means to 'do good'.

Related to the ethical considerations of 'good for whom' are questions about 'good for business'. Among participants assigned the developer role there was a repeating pattern of explanations that considered the '*success of the app or company*', to be the primary motivation for their selection of advice type to offer. This theme crossed all scenarios and both advice types, with subtle differences. For participants selecting Adoption-Directed advice the focus was on attracting a high number of downloads and '*large initial user base*', so that the app would quickly become a viable business proposition. For participants selecting Goal-Directed advice the focus was more on '*user retention and satisfaction*', on minimizing bad reviews and increasing good ones, so that the business was viable in a more ongoing sense. Examples include:

"If you are looking to generate money and a large user base with the app, then I think you would want to drive more traffic to your app. In that case, you want the option where 90% of people would participate." (*Making healthy lifestyle choices, developer, 90-15, Adoption-Directed advice*)

"Because I'd want the app to earn a good reputation by giving correct advice" (*Route planning, developer, 90-15, Goal-Directed advice*)

6.5 Experimental design choices

The choices made in designing any study are reflected in its findings. In this study, the choice between Goal-Directed and Adoption-Directed advice was intentionally presented as a one-shot binary choice, following a convention familiar in choice experiments undertaken by economists, e.g. [6, 31, 45]. We also asked participants to adopt one of three different roles when considering their selection of Goal-Directed or Adoption-Directed advice, and we selected four scenarios in which to probe participants' preferences. While similar techniques have been widely used in prior research, we should not exclude the possibility that differences observed are a function of the experiment itself. In order to gain more nuance to our understanding, and to more properly distinguish between possible experimental effects and deeper preferences, future research should consider longitudinal studies and qualitative methods. This would support investigation into if and how design choices can influence adoption rates for counter-intuitive or challenging advice, and

therefore reduce the dilemma's impact. These might be considered in the context of health, where relationships with advice givers are likely to be ongoing, and where short and long-term goals interact. To better understand the dilemma, future research should also consider including additional conditions in which Goal-Directed and Adoption-Directed advice align. Adding conditions in which the probability of individual goal attainment and the probability that the advice will be adopted are both high, and conversely where they are both low, will help to further pick apart peoples' preferences.

6.5.1 Testing for understanding the independence of advice type probabilities. To be confident that we only included data from participants who understood the trade-offs that the dilemma poses, we included a test in a second domain. Failure of this test excluded the responses of 444 participants. While we consider this the correct approach for making an initial study into the advice dilemma we identify, we also acknowledge the impact of this choice. For this reason we compare the advice type preferences of this set of 444 participants with those of the 817 participants included in our full analysis. In this comparison, we first see an overall reduction in the preference for Goal-Directed advice over Adoption-Directed advice: 65.9% (290/444) against 80.5% (658/817). A chi-square test of independence between these two samples shows that these differences are significant ($\chi^2=32.23$, $p<0.00$). When broken down by domain scenario, participants in each group also indicated a reduced preference towards the Goal-Directed advice. In three of the four scenarios this difference was significant. In the 'investment planning' scenario the difference was 60.83 % against 80.64% ($\chi^2 = 13.50$, $p < 0.00$); in the 'training for a 5k run' scenario it was 54% against 74.45% ($\chi^2 = 11.40$, $p < 0.00$); and in the 'route planning' scenario it was 76.56% against 89.64% ($\chi^2=10.5$, $p=0.001$).

Despite these differences, our initial results are robust and the findings remain largely the same with or without the exclusion. However, they do indicate that there is likely more nuance to the preferences people might have in actual use than is expressed in this initial experiment. We believe that this indicates the potential importance of numeracy skills on participants' selections, and in understanding the trade-offs the dilemma introduces. It also further indicates that research is needed into the way the dilemma manifests in the wild, where users may be exposed to repeated ongoing advice, making longitudinal decisions where short-term and long-term motivations and goals interact, and where issues of low numeracy and low engagement may be common.

6.6 Algorithmic Advice Dilemmas: Future Research

The experimental design of this study, i.e. a one-shot, discreet choice between advice types, was chosen in order to isolate the particular cases where advice might appear counter-intuitive or challenging, and echoes a tradition of similar experimental studies in economics. However, in practice this advice dilemma may also appear in cases where users have more complex, ongoing relationships with applications and recommender systems that involve use cases of repeated or adjusted advice [88], often based on users' dynamic preferences and previous choices [51, 65], and other behavioral patterns such as in sequence-aware recommender systems [73, 74].

Our quantitative findings show that participants shown the 'route planning' scenario, the scenario we consider most likely to be a one-shot decision, were also significantly more likely to select Goal-Directed advice. We also see from participants' response to the choice rationale question, that it was those shown the 'making healthier lifestyle choices' scenario who most often mention task complexity, regardless of whether they selected a preference for Goal-Directed advice or Adoption-Directed advice. This is the scenario we consider most open ended and likely to present in a situation of regular ongoing interactions between a user and a system offering algorithmic advice. Taken in combination, these two findings strongly suggest that in the context of recommender systems, responses to the dilemma will need to be more subtle and nuanced. What a user prefers

may vary across interactions, and may also vary through the temporal unfolding of interactions. Their preferences before advice is offered may vary from their preferences at the time of offering, they may vary again following success or failure in adhering to that advice, and they may vary yet again following information about the choices of other users. Future research, building on work at the intersection of CSCW and recommender systems [23, 53, 89], would therefore be needed to address the design dilemma outlined here in the context of repeated advice and collaborative filtering [27]. For example, should algorithmic tools prioritize a sub-optimal ‘good’ advice over optimal ‘best’ advice if the sub-optimal advice is consistent with the choices of users in the network of the focal user, and therefore more likely to be followed? Such questions should be considered not only in terms of their effectiveness in driving users’ choices, but importantly, in relation the ethical aspects of recommender systems and their design (e.g [63, 71]).

6.7 Design Considerations

In this paper we probe a dilemma for designers of algorithmic advice systems. Here we discuss design considerations for these circumstances. These we consider complementary to existing guidelines, e.g. for designing social recommender systems [4, 41].

Once identified, the dilemma between Goal-Directed and Adoption-Directed advice should be surfaced early on in the design process. Human-centered design methods advocate for iterative cycles of contextual inquiry, prototyping, and evaluation; and at each stage we would argue the need for comparative user studies, both qualitative and quantitative, to explore the dilemma as it plays out in-situ. This would allow not only comparison between different advice types, but also inquiry into specific ways they might be dynamically explored and compared by users; e.g. in situations such as route planning where the alternative recommendations may be presented simultaneously, but where designers must choose which advice to suggest as initially preferred. Our findings suggest not only that preferences may vary according to domain and scenario, but also that motivations underlying these preferences may vary too.

While our findings may indicate participants’ clear preference for Goal-Directed advice, a substantial minority of participants selected a preference for Adoption-Directed advice. In practice, the advice design dilemma is often likely to present in complex situations where users have ongoing longitudinal interactions with applications offering advice. The variation we see in the strength of preference between the different domain scenarios, and the explanations participants gave for making these selections, across and within these scenarios, suggest that designers should investigate adaptive approaches to presenting algorithmic advice. At a simple level, this might involve making advice type a dynamic and/or configurable option, either when the user is completing initial setup or in use on a case by case situational basis. However, designers might also consider options that draw on prior research that inspired this work. For example, if we know that users have a tendency to discount good advice [36, 60] and display overconfidence in their own abilities [13] because of a bias towards self-related, egocentric information [26, 30, 54, 55], can the choice of advice type be combined with challenges based on predicted goal success and adoption rates? and can the way this is presented to users guide these users towards more effective decision-making? In such a case, we can imagine designing an application that uses machine learning to predict the likelihood of the advice dilemma, e.g. in advice relating to lifestyle changes, and which then adapts the advice offered dynamically, based first on data from aggregated prior use and then increasingly integrating this with the users’ own responses. We can also imagine a potential user who receives advice, such as to quit smoking or to severely reduce salt or sugar consumption, but discounts the advice because they may be overconfident in their capacity to avoid the consequences, or because they are biased towards their current feelings of wellbeing. In this situation, the application might start - upon user consent - by varying the advice, at times using low adoption rates associated

with Goal-Directed advice as an egocentric challenge to the user, and at other times using the less challenging but still beneficial goals associated with Adoption-Directed advice to provide a confidence supporting reward. Over time the application might blend its wider understanding of what represents Goal-Directed or Adoption-Directed advice, which is based on data aggregated from many users, with a more nuanced understanding of what these might mean for the particular user in question, and offer advice that responds to the dynamics of a particular situation in a more fine-grained way.

The example of applications that provide users with healthcare advice illustrates the considerations necessary for design and development in the medical domain. Digital therapeutics systems [66], for managing chronic diseases such as diabetes, are flourishing, with healthcare clinicians now prescribing apps to patients as supplements to their treatment regimens. In this rapidly evolving situation, designers play an important role in helping shape the messaging that clinicians leverage through new digital therapeutic interactions. Given a choice, does a physician prescribe the diabetes app that is most challenging to follow but has the best clinical outcomes when adhered to; or one that is more easily adopted by a larger number of patients, albeit with a lower effect size on those that follow its advice? What factors should influence this decision? And how might it be affected by the prescriber's inherent biases? Is there a role for advanced analytics, based on social cues from historic use, in helping to personalize the advice these recommender systems offer, so that the clinician can promote adoption directed messages in patients who are likely to struggle with the more challenging path and goal directed advice for those who demonstrate a higher probability of completion? For the clinician, these trade-offs are not new. However, sharing the way these choices influence patients with designers or developers is. This research indicates that new partnerships and collaborations between designers and healthcare clinicians are urgently needed to examine these trade-offs and develop best practices.

7 CONCLUSION

In this paper we have made an initial step towards better understanding the advice design dilemma that arises when advice most likely to maximize an individual's chances of achieving their goal is not the same as advice most likely to maximize the number of people who will adopt it. We studied participants' preferences when selecting between Goal-Directed and Adoption-Directed advice online in an experimental study spanning four domain scenarios, three user types, and two variations in the gap between goal attainment and advice adoption probabilities. We found an average preference for advice that favors individual goal attainment over higher user adoption rates, albeit with some variation across advice domains. The implications of these findings are both practical and theoretical. From a practical perspective, choosing between presenting advice that may be best in terms of individual goal attainment but which will reach few users, or advice that is merely good but will be adopted by many, is likely to be an increasingly common dilemma for designers to face in algorithmic advice design scenarios. From a theoretical perspective, the preferences shown for Goal-Directed advice (i.e. advice best for individual goal attainment that will be adopted by few users), and the explanations revealed in participants' survey comments, echo responses to moral dilemmas, e.g. [15, 82], where people eschew making decisions that are sub-optimal at the individual level, even if they are optimal at the population level. With the rapid growth in applications driven by algorithmic advice based on the social information derived from aggregated prior use, HCI and CSCW researchers and practitioners are likely to increasingly engage with these fundamental questions. Our findings will help inform this conversation.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation award #1928614.

REFERENCES

- [1] Gediminas Adomavicius, Jesse Bockstedt, Shawn Curley, and Jingjing Zhang. 2019. Reducing Recommender Systems Biases: An Investigation of Rating Display Designs. *Forthcoming, MIS Quarterly* (2019), 19–18.
- [2] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering* 6 (2005), 734–749.
- [3] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 217–253.
- [4] Ofer Arazy, Nanda Kumar, and Bracha Shapira. 2010. A theory-driven design framework for social recommender systems. *Journal of the Association for Information Systems* 11, 9 (2010), 455.
- [5] Amos Azaria, Sarit Kraus, Claudia V Goldman, and Omer Tsimhoni. 2014. Advice provision for energy saving in automobile climate control systems. In *Twenty-Sixth IAAI Conference*.
- [6] Nick Bansback, John Brazier, Aki Tsuchiya, and Aslam Anis. 2012. Using a discrete choice experiment to estimate health state utility values. *Journal of health economics* 31, 1 (2012), 306–318.
- [7] Nicholas C Barberis. 2013. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives* 27, 1 (2013), 173–96.
- [8] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 675–684.
- [9] Shlomo Benartzi and Richard Thaler. 2007. Heuristics and biases in retirement savings behavior. *Journal of Economic perspectives* 21, 3 (2007), 81–104.
- [10] Shlomo Benartzi and Richard H Thaler. 1999. Risk aversion or myopia? Choices in repeated gambles and retirement investments. *Management science* 45, 3 (1999), 364–381.
- [11] Daniel Berdichevsky and Erik Neuenschwander. 1999. Toward an ethics of persuasive technology. *Commun. ACM* 42, 5 (1999), 51–58.
- [12] Eli Blevis. 2007. Sustainable interaction design: invention & disposal, renewal & reuse. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 503–512.
- [13] Richard A Block and David R Harper. 1991. Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational behavior and human decision processes* 49, 2 (1991), 188–207.
- [14] Silvia Bonaccio and Reeshad S Dalal. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational behavior and human decision processes* 101, 2 (2006), 127–151.
- [15] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352, 6293 (2016), 1573–1576.
- [16] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [17] Erik Brynjolfsson, Avinash Collis, and Felix Eggers. 2019. Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences* 116, 15 (2019), 7250–7255.
- [18] Richard Buchanan. 1985. Declaration by design: Rhetoric, argument, and demonstration in design practice. *Design issues* (1985), 4–22.
- [19] Jürgen Buder and Christina Schwind. 2012. Learning with personalized recommender systems: A psychological view. *Computers in Human Behavior* 28, 1 (2012), 207–216.
- [20] Leonardo Bursztyn, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica* 82, 4 (2014), 1273–1301.
- [21] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 503.
- [22] Mark Cartwright, Graham Dove, Ana Elisa Méndez Méndez, Juan P Bello, and Oded Nov. 2019. Crowdsourcing multi-label audio annotation tasks with citizen scientists. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 292.
- [23] Shuo Chang, F Maxwell Harper, and Loren Terveen. 2015. Using groups of items to bootstrap new users in recommender systems. In *CSCW 2015: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1258–1269.
- [24] Shruthi Sai Chivukula, Colin M Gray, and Jason A Brier. 2019. Analyzing Value Discovery in Design Decisions Through Ethicography. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 77.
- [25] Robert B Cialdini and Robert B Cialdini. 1993. Influence: The psychology of persuasion. (1993).
- [26] Russell W Clement and Joachim Krueger. 2000. The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology* 39, 2 (2000), 279–299.

- [27] Nediya Daskalova, Bongshin Lee, Jeff Huang, Chester Ni, and Jessica Lundin. 2018. Investigating the effectiveness of cohort-based sleep recommendations. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–19.
- [28] Stefano DellaVigna. 2009. Psychology and economics: Evidence from the field. *Journal of Economic literature* 47, 2 (2009), 315–72.
- [29] Carl DiSalvo, Phoebe Sengers, and Hrönn Brynjarsdóttir. 2010. Mapping the landscape of sustainable HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1975–1984.
- [30] David Dunning and Andrew F Hayes. 1996. Evidence for egocentric comparison in social judgment. *Journal of personality and social psychology* 71, 2 (1996), 213.
- [31] Michael Egesdal, Zhenyu Lai, and Che-Lin Su. 2015. Estimating dynamic discrete-choice games of incomplete information. *Quantitative Economics* 6, 3 (2015), 567–597.
- [32] Gunther Eysenbach, John Powell, Marina Englesakis, Carlos Rizo, and Anita Stern. 2004. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *Bmj* 328, 7449 (2004), 1166.
- [33] Avi Fleischer, Alan D Mead, and Jialin Huang. 2015. Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology* 8, 2 (2015), 196–202.
- [34] Batya Friedman and Peter H Kahn Jr. 2003. Human values, ethics, and design. *The human-computer interaction handbook* (2003), 1177–1201.
- [35] Jeff Galloway. [n. d.]. Training: Run, Walk, Run. Retrieved August 28, 2019 from <http://www.jeffgalloway.com/training/run-walk/>
- [36] Paul Goodwin and Robert Fildes. 1999. Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making* 12, 1 (1999), 37–53.
- [37] Junius Gunaratne and Oded Nov. 2015. Influencing retirement saving behavior with expert advice and social comparison as persuasive techniques. In *International Conference on Persuasive Technology*. Springer, 205–216.
- [38] Junius Gunaratne and Oded Nov. 2015. Informing and improving retirement saving performance using behavioral economics theory-driven user interfaces. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, 917–920.
- [39] Junius Gunaratne, Lior Zalmanson, and Oded Nov. 2018. The Persuasive Power of Algorithmic and Crowdsourced Advice. *Journal of Management Information Systems* 35, 4 (2018), 1092–1120.
- [40] Jacob Hacker and Ann O’Leary. 2012. *Shared responsibility, shared risk: government, markets and social policy in the twenty-first century*. OUP USA.
- [41] Jason L Harman, John O’Donovan, Tarek Abdelzaher, and Cleotilde Gonzalez. 2014. Dynamics of human trust in recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems*. 305–308.
- [42] Björn Hartmann, Daniel MacDougall, Joel Brandt, and Scott R Klemmer. 2010. What would other programmers do: suggesting solutions to error messages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1019–1028.
- [43] David J Hauser and Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods* 48, 1 (2016), 400–407.
- [44] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.
- [45] Matt Horne, Mark Jaccard, and Ken Tiedemann. 2005. Improving behavioral realism in hybrid energy-economy models using discrete choice studies of personal transportation decisions. *Energy Economics* 27, 1 (2005), 59–77.
- [46] Dietmar Jannach, Paul Resnick, Alexander Tuzhilin, and Markus Zanker. 2016. Recommender systems—Beyond matrix completion. *Commun. ACM* 59, 11 (2016), 94–102.
- [47] Eva Jonas, Stefan Schulz-Hardt, Dieter Frey, and Norman Thelen. 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. *Journal of personality and social psychology* 80, 4 (2001), 557.
- [48] Garth S Jowett and Victoria O’donnell. 2018. *Propaganda & persuasion*. Sage Publications.
- [49] Daniel Kahneman and Richard H Thaler. 2006. Anomalies: Utility maximization and experienced utility. *Journal of Economic Perspectives* 20, 1 (2006), 221–234.
- [50] Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47, 2 (1979), 363–391.
- [51] Komal Kapoor, Vikas Kumar, Loren Terveen, Joseph A Konstan, and Paul Schrater. 2015. "I like to explore sometimes" Adapting to Dynamic User Novelty Preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 19–26.
- [52] Flavius Kehr, Marc Hassenzahl, Matthias Laschke, and Sarah Diefenbach. 2012. A transformational product to improve self-control strength: the chocolate machine. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems. ACM, 689–694.

- [53] Sami Koivunen, Thomas Olsson, Ekaterina Olshannikova, and Aki Lindberg. 2019. Understanding Decision-Making in Recruitment: Opportunities and Challenges for Information Technology. *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (2019), 1–22.
- [54] Joachim Krueger and David Stanke. 2001. The role of self-referent and other-referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin* 27, 7 (2001), 878–888.
- [55] Justin Kruger. 1999. Lake Wobegon be gone! The below-average effect and the egocentric nature of comparative ability judgments. *Journal of personality and social psychology* 77, 2 (1999), 221.
- [56] Susan Lacke. 2017. 3 Run/Walk Interval Workouts To Incorporate Into Your Training. Retrieved August 28, 2019 from https://www.womensrunning.com/2017/09/training/run-walk-interval-workouts_81066
- [57] Richard N Landers and Tara S Behrend. 2015. An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology* 8, 2 (2015), 142–164.
- [58] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2011. Mining behavioral economics to design persuasive technology for healthy choices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 325–334.
- [59] Priel Levy and David Sarne. 2016. Intelligent advice provisioning for repeated interaction. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- [60] Joa Sang Lim and Marcus O'Connor. 1995. Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making* 8, 3 (1995), 149–168.
- [61] Hazel R Markus and Shinobu Kitayama. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review* 98, 2 (1991), 224.
- [62] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.
- [63] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. 2020. Recommender systems and their ethical challenges. *AI & SOCIETY* (2020), 1–11.
- [64] Stanley Milgram. 1963. Behavioral study of obedience. *The Journal of abnormal and social psychology* 67, 4 (1963), 371.
- [65] Joshua L Moore, Shuo Chen, Douglas Turnbull, and Thorsten Joachims. 2013. Taste Over Time: The Temporal Dynamics of User Preferences. In *ISMIR*. Citeseer, 401–406.
- [66] Camille Nebeker, John Torous, and Rebecca J Bartlett Ellis. 2019. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC medicine* 17, 1 (2019), 137.
- [67] Duyen T Nguyen, Laura A Dabbish, and Sara Kiesler. 2015. The perverse effects of social transparency on online advice taking. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 207–217.
- [68] Oded Nov and Ofer Arazy. 2013. Personality-targeted design: theory, experimental procedure, and preliminary results. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM, 977–984.
- [69] U.S. Department of Health, Human Services, and U.S. Department of Agriculture. 2015. Dietary Guidelines for Americans. Retrieved August 28, 2019 from <https://health.gov/dietaryguidelines/2015/guidelines/>
- [70] Dilek Önköl, Paul Goodwin, Mary Thomson, Sinan Gönöl, and Andrew Pollock. 2009. The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making* 22, 4 (2009), 390–409.
- [71] Dimitris Paraschakis. 2016. Recommender systems from an industrial and ethical perspective. In *Proceedings of the 10th ACM conference on recommender systems*. 463–466.
- [72] 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 *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 423–432.
- [73] Massimo Quadrana and Paolo Cremonesi. 2018. Sequence-aware recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 539–540.
- [74] Massimo Quadrana, Paolo Cremonesi, and Dietmar Jannach. 2018. Sequence-aware recommender systems. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–36.
- [75] Johan Redström. 2006. Persuasive design: Fringes and foundations. In *International Conference on Persuasive Technology*. Springer, 112–122.
- [76] Briane Paul V Samson and Yasuyuki Sumi. 2019. Exploring Factors that Influence Connected Drivers to (Not) Use or Follow Recommended Optimal Routes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 371.
- [77] Christina Schwind, Jürgen Buder, and Friedrich W Hesse. 2011. I will do it, but i don't like it: user reactions to preference-inconsistent recommendations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 349–352.

- [78] Hersh M Shefrin and Richard H Thaler. 1988. The behavioral life-cycle hypothesis. *Economic inquiry* 26, 4 (1988), 609–643.
- [79] Kirsten Swearingen and Rashmi Sinha. 2001. Beyond algorithms: An HCI perspective on recommender systems. In *ACM SIGIR 2001 workshop on recommender systems*, Vol. 13. Citeseer, 1–11.
- [80] Richard H Thaler. 2018. From cashews to nudges: the evolution of behavioral economics. *American Economic Review* 108, 6 (2018), 1265–87.
- [81] Richard H Thaler and Shlomo Benartzi. 2007. The behavioral economics of retirement savings behavior. (2007).
- [82] Judith Jarvis Thomson. 1976. Killing, letting die, and the trolley problem. *The Monist* 59, 2 (1976), 204–217.
- [83] Harry C Triandis, Christopher McCusker, and C Harry Hui. 1990. Multimethod probes of individualism and collectivism. *Journal of personality and social psychology* 59, 5 (1990), 1006.
- [84] Chun-Hua Tsai and Peter Brusilovsky. 2017. Providing control and transparency in a social recommender system for academic conferences. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. 313–317.
- [85] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *science* 185, 4157 (1974), 1124–1131.
- [86] Tom R Tyler and E Allan Lind. 1992. A relational model of authority in groups. In *Advances in experimental social psychology*. Vol. 25. Elsevier, 115–191.
- [87] Weiquan Wang and Izak Benbasat. 2007. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* 23, 4 (2007), 217–246.
- [88] Xiang Wang, Xiangnan He, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. 2018. Tem: Tree-enhanced embedding model for explainable recommendation. In *Proceedings of the 2018 World Wide Web Conference*. 1543–1552.
- [89] Fengli Xu, Zhenyu Han, Jinghua Piao, and Yong Li. 2019. "I Think You'll Like It" Modelling the Online Purchase Behavior in Social E-commerce. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
- [90] Ilan Yaniv and Eli Kleinberger. 2000. Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes* 83, 2 (2000), 260–281.
- [91] Rodrigo Zenun Franco. 2017. Online Recommender System for Personalized Nutrition Advice. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 411–415.

Received January 2020; revised June 2020; accepted July 2020