The Impact of Placebic Explanations on Trust in Intelligent Systems

Malin Eiband, Daniel Buschek, Alexander Kremer, Heinrich Hussmann
LMU Munich, Munich, Germany
malin.eiband,daniel.buschek,hussmann@ifi.lmu.de,alexander.kremer@campus.lmu.de

ABSTRACT

Work in social psychology on interpersonal interaction [5] has demonstrated that people are more likely to comply to a request if they are presented with a justification – even if this justification conveys no information. In the light of the many calls for explaining reasoning of interactive intelligent systems to users, we investigate whether this effect holds true for human-computer interaction. Using a prototype of a nutrition recommender, we conducted a lab study (N=30) between three groups (*no explanation*, *placebic explanation*, and *real explanation*). Our results indicate that placebic explanations for algorithmic decision-making may indeed invoke perceived levels of trust similar to real explanations. We discuss how placebic explanations could be considered in future work.

CCS CONCEPTS

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

KEYWORDS

Explainability; explanations; transparency; intelligent systems; XAI.

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312787

ACM Reference Format:
Malin Eiband, Daniel Buschek, Alexander Kremer, Heinrich Hussmann. 2019. The Impact of Placebic Explanations on Trust in Intelligent Systems. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3290607.3312787

As everyone in your family knows, you are very tech-savvy and always familiar with the latest trends. Your mother has been interested in a healthier diet for a long time and has been thinking about a nutrition plan that would allow her to lose some weight without being too restrictive. She asks you if there is a good app that could help her with her goal. By chance, a few days ago you heard about a new app that can help her do just that. To enable you to create a personalised plan for her, you have the following details about her and her dietary preferences:

Sex: Female, Age: 47, Height: 1,68m, Weight: 78kg. She works full-time as a nurse in a local hospital. She would like to lose between 4kg and 6kg in a 3-month period. She likes salad the least and if it is possible, she would like to eat something else as long as it does not affect the result too much.

Now, with the help of her information, put together a plan for your mum.

Sidebar 1: Scenario given to the participants in the study.

INTRODUCTION & BACKGROUND

Intelligent systems, that is, systems employing machine learning techniques, are now an integral part of many applications that we use on a daily basis. Yet, the black-box nature of many machine learning algorithms violates usability principles established in human-computer interaction (HCI) like easy error correction and predictability of system output [1, 3]. *Explanation* of algorithmic decision-making is therefore widely called for as a way to making such systems transparent and comprehensible (e.g. [7]).

Importantly for this work, it has been noted that the process of explanation is subject to complex cognitive and social processes and biases in human communication [6]. In particular, we investigate an effect observed in a social psychology experiment by Langer et al. [5]. They asked 120 participants approaching the copy machine of a library to let another person (one of the researchers) go first and observed if they complied to the request or not. The dependent variables were the *effort* compliance involved (*small* or *big*, i.e. letting the researcher copy five or 20 pages) and the *justification* given: The researcher either (1) provided *no explanation*: "May I use the xerox machine?", (2) gave a *"placebic" explanation* that did not contain any information: "May I use the xerox machine, because I have to make copies?", or (3) used a *"real" explanation*: "May I use the xerox machine, because I'm in a rush?".

The results of their study reveal that people are more likely to comply to a request if presented with a justification – *even if this justification conveys no information*. In fact, when the involved effort was small, the compliance rate of the placebic-explanation and real-explanation group were almost identical, 93 % and 94 %, and significantly higher than that of the no-explanation group (60 %).

Although the study is located in a social setting, it has been echoed in the literature on intelligent systems. Weller [8] and Zerilli et al. [9] both draw attention to possible pitfalls of Langer et al.'s results with regard to explainability and transparency. As Weller puts it, "a possible worry is that a deployer might provide an empty explanation as a psychological tool to soothe users". In this work, we therefore seek answer to the question:

> *Do placebic explanations invoke a similar level of trust in an intelligent system as real explanations?*

To the best of our knowledge, we present the first user study and results on this question.

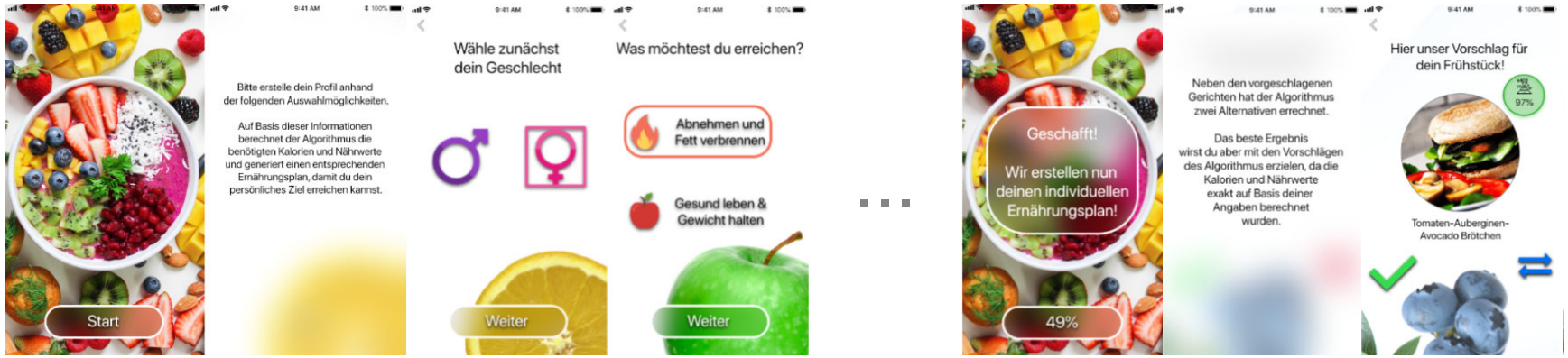

Figure 1: Exemplary screenshots as illustration of the prototype (in German) used in the study in the real-explanations version.

USER STUDY

To transfer Langer et al.'s small-effort condition to the context of intelligent systems, we focused on a low-risk application people encounter in their everyday life for our study. Namely, we decided to use a nutrition recommendation system which suggests personalised meals to support users lose weight.

Study Design

In line with Langer et al.'s work, we chose a between-groups design for our study with one group per type of explanation (*no explanation*, *placebic explanation*, *real explanation*). We assessed the consequent perception of trust in the app and its recommendations through a questionnaire. The questions were taken from work by Corritore et al. [2] and complemented by us through three questions on the perceived understanding of the algorithmic decision-making. All questions were assessed on 5-point Likert scales (1=strongly disagree to 5=strongly agree). The no-explanation group used the prototype without any explanation of algorithmic decision-making. The explanations given to the placebic-explanation and real-explanation groups, respectively, can be found in Table 2 (translated to English). Placebic explanations were phrased so as to semantically introduce a justification with "because/since/so that ..." (similar to Langer et al.'s question phrasing), but to not convey more information about the algorithm than could be inferred from the study scenario described in the next sections. The real explanations included details about the (apparent) algorithmic decision-making as well as the system certainty for a particular recommendation.

Placebic Explanation	Real Explanation
We need these details because they are necessary for the algorithm.	Based on this information, the algorithm calculates the need for calories and nutritional values and generates a corresponding nutrition plan so that you can reach your personal goal.
Please enter your age, weight and height because the algorithm takes them into account.	The details about your age, weight and height are necessary because the algorithm uses them to calculate several numbers. These numbers, such as the minimal calorie intake, the BMI and nutritional values will then be used to create your personal nutrition plan.
Please indicate your diet goal so that the algorithm can adjust your plan accordingly.	Please indicate your diet goal so that the algorithm can take it into account for your nutrition plan and respective nutritional values.
The algorithm has calculated one recommendation and two alternatives, but you will reach the best result with the recommendation since it was calculated by the algorithm.	The algorithm has calculated one recommendation and two alternatives, but you will reach the best result with the recommendation since the number of calories and nutritional values has been calculated based on your personal details.

Sidebar 2: Placebic and real explanations used in our prototype.

Prototype

The prototype we used was realised as a click-dummy mobile app for proof of concept. To allow for a controlled experiment and comparability between the three groups, we did not implement the system, but instead simulated personalised recommendations. The design of the prototype was inspired by current popular nutrition recommender apps. Figure 1 shows exemplary screenshots for illustration.

The prototype consisted of two parts: an onboarding part where details about a user (e.g. age, weight, dietary goal) are entered, and the subsequent meal recommendations for breakfast, lunch and dinner (each with two alternatives to choose from). Depending on the group condition, the prototype either included no explanations, placebic explanations or real explanations in the onboarding as well as the recommendation part.

Participants & Procedure

We recruited 30 participants aged between 22 and 43 (median 28 years), 13 females and 17 males, and invited them to our lab. We assigned them randomly to one of the three groups so that each group eventually consisted of 10 people. All participants indicated an IT background and were German.

After filling out a consent form, participants were presented with the scenario and associated task shown in Sidebar 1 (translated to English). We chose to let participants use the app in the name of someone else (their "mother") and not for themselves to exclude any possible impact of personal food preferences. Moreover, the scenario was designed so as to leave room for a certain ambiguity with regard to the system's decision (the mother's dislike of salad).

Participants were then given a mobile phone with the respective prototype version to complete the task. They were first led through the onboarding process where they filled in the mother's details. After that, they received the app's recommendation and two alternatives to choose from for each breakfast, lunch and dinner to create a personalised nutrition plan. For lunch, the app was set up so as to show a salad as the recommended meal. Participants had the opportunity to scroll through the meals to get an overview of their options before making their choice. At the end, we asked the participants about the reasons for their choices and then handed out the questionnaire. The study took 10-15 minutes in total to complete.

RESULTS

We give a descriptive account of the answers we received due to the small sample size per group.

Perceived Trust in the App and Explanation Scope

Figure 2 compares the median of the answers to the questionnaire for the three groups, Figure 3 to 5 show the answers to the questionnaire within each group. The first three questions target the

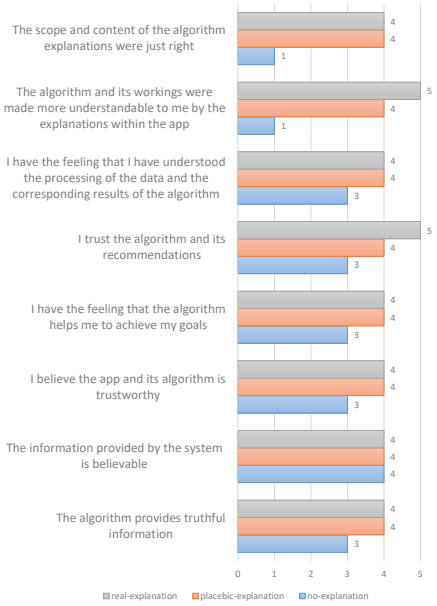

Figure 2: Median of the answers to the questionnaire for the three groups.

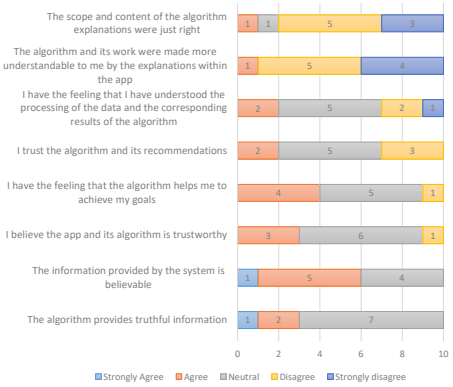

Figure 3: No-explanation group: Results of the questionnaire (absolute numbers).

perceived understanding of the algorithmic decision-making, the rest captures the perceived trust in the app and its recommendations [2].

When comparing the medians between groups, they indicate a difference between the no-explanation group on the one side and the placebic- and real-explanation groups on the other. Notably, the median of the answers to the trust questions is, apart from one exception, always larger in the placebic- and real-explanation groups compared to the no-explanation group. Although the numbers are small, we thus observe a tendency for a similar effect as found by Langer et al. in their work: Placebic and real explanations both seem to lead to more perceived trust in the system. In particular, placebic explanations may indeed invoke similar levels of perceived trust as real explanations. Moreover, it is interesting to see that both the understandability as well as the scope and content of the explanations were perceived as being similarly sufficient in both the placebic- and real-explanation group.

Choice of Meals

We were interested in seeing if the group condition had an influence on the selection of meals, in particular the salad which conflicts with the mother's food preferences.

None of the 10 participants in the no-explanation group opted for the salad although it was displayed as the recommended meal for lunch. All 10 stated that they had chosen a different meal due to the mother's personal dislike of salad.

In the placebic-explanation group, 4 participants selected the salad. All of these 4 participants in the placebic-explanation group said that they wanted to achieve the best possible result for the mother and therefore attached greater importance to the recommendation of the algorithm than to the mother's preferences. 3 participants said that while they were aware that the algorithm promised a better result with the salad, they did not want to disregard the mother's preferences. The remaining 3 participants found the information insufficient to disregard the mother's preferences.

In the real information group 4 participants opted for the salad, too. They said that they had followed the recommendation because of the high system certainty. On the other hand, 5 participants stated that they deliberately decided not to choose the salad despite the high system certainty because the difference to the given alternatives seemed not big enough to neglect the mother's preferences. The remaining participant did not convey any reason for her choice.

DISCUSSION & CONCLUSION

Our study's sample is limited in size and diversity. Thus it may not represent the general population. Moreover, the influence of the system domain (e.g. recommending nutrition vs financial products) was not validated in our study and the presented results may not generalise to other systems. Future work could extend the sample, for instance, by running the study as an online survey, and investigate other contexts.

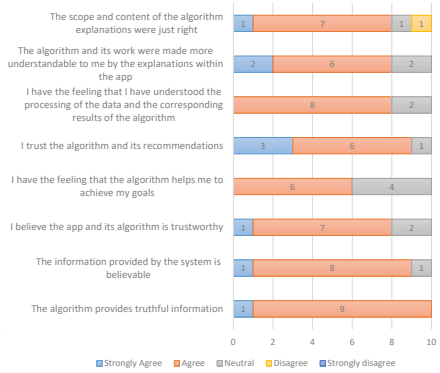

Figure 4: Placebic-explanation group: Results of the questionnaire (absolute numbers).

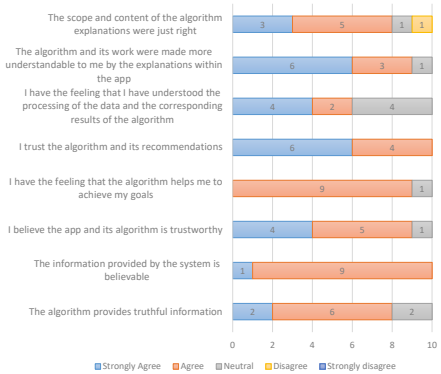

Figure 5: Real-explanation group: Results of the questionnaire (absolute numbers).

Nevertheless, our results indicate that placebic explanations for algorithmic decision-making may invoke perceived levels of trust similar to real explanations and thus motivate large scale investigation of the psychological effects of placebic explanations. This placebo effect seems potentially worrisome if used to "deceive" users in a sense comparable to "dark" UX patterns [4]. On the other hand, placebic explanations might play a useful role, for example, as a placeholder/default until enough information for a real explanation has been collected (e.g. for a new user in a personalised system). Moreover, as one consequence for HCI research, future work on explanations for intelligent systems might consider using a placebic explanation as a baseline, not (only) a baseline without any explanations at all. Only then may observed effecs be attributed to the actual explanation content, and not merely to the psychological effect of a placebic explanation.

In future work, we plan to deepen our investigations: For example, we deem it worthwhile to test placebic vs real explanations possibly separated from a particular prototype. Moreover, it would be interesting to see if placebic explanation also has an effect in a big-effort condition in which users have to make high-risk decisions such as in medical systems.

REFERENCES

[1] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35, 4 (December 2014), 105. https://doi.org/10.1609/aimag.v35i4.2513

[2] Cynthia L. Corritore, Robert P. Marble, Susan Wiedenbeck, Beverly Kracher, and Ashwin Chandran. 2005. Measuring Online Trust of Websites: Credibility, Perceived Ease of Use, and Risk. In *A Conference on a Human Scale. 11th Americas Conference on Information Systems, AMCIS 2005, Omaha, Nebraska, USA, August 11-14, 2005*. 370. http://aisel.aisnet.org/amcis2005/370

[3] John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Transactions on Interactive Intelligent Systems* 8, 2, Article 8 (June 2018), 37 pages. https://doi.org/10.1145/3185517

[4] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 534, 14 pages. https://doi.org/10.1145/3173574.3174108

[5] Ellen J. Langer, Arthur Blank, and Benzion Chanowitz. 1978. The Mindlessness of Ostensibly Thoughtful Action: The Role of "Placebic" Information in Interpersonal Interaction. *Journal of Personality and Social Psychology* 36, 6 (1978), 635–642. https://doi.org/10.1037/0022-3514.36.6.635

[6] Tim Miller. 2017. Explanation in Artificial Intelligence: Insights from the Social Sciences. *CoRR* abs/1706.07269 (2017). arXiv:1706.07269 http://arxiv.org/abs/1706.07269

[7] Don Monroe. 2018. AI, Explain Yourself. *Commun. ACM* 61, 11 (Oct. 2018), 11–13. https://doi.org/10.1145/3276742

[8] Adrian Weller. 2017. Challenges for Transparency. *CoRR* abs/1708.01870 (2017). arXiv:1708.01870 http://arxiv.org/abs/1708.01870

[9] John Zerilli, Alistair Knott, James Maclaurin, and Colin Gavaghan. 2018. Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology* (2018). https://doi.org/10.1007/s13347-018-0330-6