Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4079–4087
Marseille, 20-25 June 2022

Corpus Design for Studying Linguistic Nudges in Human-Computer Spoken Interactions

Natalia Kalashnikova[*†], **Serge Pajak**[¶†], **Fabrice Le Guel**[¶†], **Ioana Vasilescu**[*]
Gemma Serrano[‡], **Laurence Devillers**[*§]
[*]LISN-CNRS, [†]Université Paris-Saclay, [¶]RITM, [‡]Collège des Bernardins, [§]Sorbonne-University
natalia.kalashnikova@limsi.fr, serge.pajak@universite-paris-saclay.fr, fabrice.le-guel@universite-paris-saclay.fr,
ioana@limsi.fr, gemma.serrano@collegedesbernardins.fr, devil@limsi.fr

Abstract

In this paper, we present the methodology of corpus design that will be used to study the comparison of influence between linguistic nudges with positive or negative influences and three conversational agents: robot, smart speaker, and human. We recruited forty-nine participants to form six groups. The conversational agents first asked the participants about their willingness to adopt five ecological habits and invest time and money in ecological problems. The participants were then asked the same questions but preceded by one linguistic nudge, with positive or negative influence. The comparison of standard deviation and mean metrics of differences between these two notes (before the nudge and after) showed that participants were mainly affected by nudges with positive influence, even though several nudges with negative influence decreased the average note. In addition, participants from all groups were willing to spend more money than time on ecological problems. In general, our experiment's early results suggest that a machine agent can influence participants to the same degree as a human agent. A better understanding of the power of influence of different conversational machines and the potential of influence of nudges of different polarities will lead to the development of ethical norms of human-computer interactions.

Keywords: linguistic nudges, corpus design, human-computer spoken interactions

1. Introduction

The concept of nudges was highlighted in Economics by Thaler and Sunstein (2008). They defined *nudges* as techniques of indirect suggestion and positive reinforcement that can influence people's choices while not being a direct restriction of their choices. Importantly, a nudge should improve social welfare but it can also work against your interests (Thaler, 2018).

In this study, a *"nudge with positive influence"* steers a participant to adopt a habit by presenting positive side of the consequences of this habit. The *"nudge with negative influence"* focuses on the negative side of a habit's outcomes, pushing a participant in the opposite direction (not adopting this habit). The terms "positive" and "negative" indicate whether information presents advantages (positive) or disadvantages (negative) of a habit.

The explicit topic of the conversation between conversational agents and participants is their attitude towards the climate issues and their ecological practices. People today constantly receive messages about environmental issues that motivate them to help the environment. When people receive information that goes against mainstream, it may require more cognitive effort for them to process it.

In the domain of spoken interactions, linguistic nudges can be used by conversational agents and connected objects to simplify people's lives. However, there are a few limits of ethical norms for a conversational agent to enter a more private zone, thereby influencing opinions or purchases. Therefore, a study investigating people's potential to be influenced by machines is needed.

In this paper, we describe the methodology which was tested during the first data collection with the aim to investigate the potential of influence by the robot, the smart-speaker, and the human agent and nudges with positive and negative information. The proposed method consists of two major steps. During the first step, we measure the global level of involvement and the willingness to adopt a specific behavior. In the second step, we present positive and negative outcomes of the five ecological habits to affect participants' willingness to adopt these habits.

We measured standard deviation and mean metrics of difference between notes after and before nudging to analyze tendencies of participants' reaction to nudges with positive and negative influences, as well as various interlocutors.

In the remainder of this paper, we introduce some related works on nudging theory and the study of nudges in linguistics (Sec. 2). We subsequently present the experimental design (Sec. 3), the analysis (Sec. 4), and the future steps for data collection and ideas of improving methodology (Sec. 5).

2. Related Work

Nudges are defined as indirect techniques that guide people's choice in a particular direction by making it easier to go into that direction, while not imposing any direct restriction on the choice set, or any typical incentives such as monetary consequences for a choice (Thaler and Sunstein, 2008). The scientists distinguish two kinds of nudges : *"nudge for good"* and *"nudge for evil"* (Sunstein, 2020). *"Nudges for good"* make a choice easier for personal or societal advantage, such

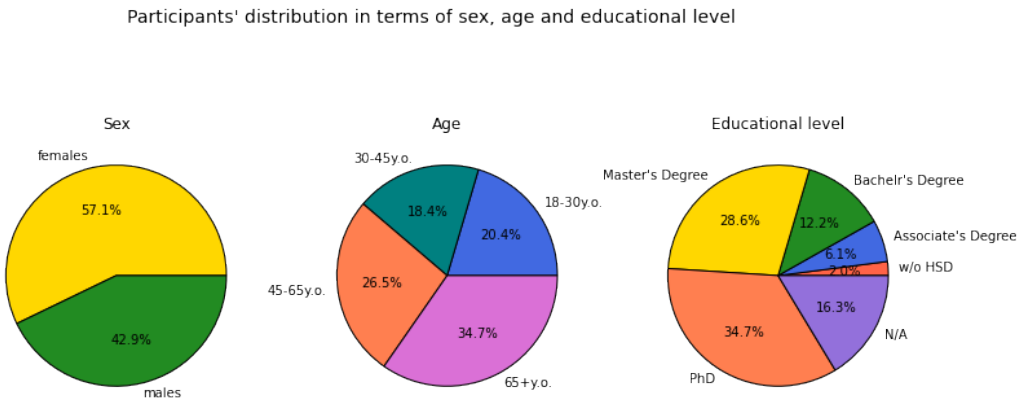

Figure 1: Participants' distribution in terms of sex, age and educational level for the recording session.

as automatic enrollment in a beneficial retirement program or an opt-out donation program, which presumes that residents of a country are willing organ donors (Etheredge, 2021). *"Nudges for evil"* serve others' interests. For example, an easy and quick check-out at e-commerce sites (Sunstein, 2020).

At the early stage of the nudging theory, *"nudges for evil"* were interchangeably used with the notion of *"sludges"*, which were determined as initiatives that encourage self-destructing behavior (Thaler, 2018). Later, sludges are mostly considered as a particular kind of nudges, which makes decision-making more difficult (Mills, 2020; Shahab and Lades, 2021) leading to *"frictions that make good decisions harder"* (Sunstein, 2020). Within these techniques, *"sludges for good"* (e.g., a waiting period) and *"sludges for evil"* (e.g., complicated form-filling) are also distinguished (Sunstein, 2020).

In this study, we used nudges to present positive and negative outcomes of ecological habits to motivate participants to reflect on their ecological behavior. We adopted the terms *"nudges with a positive influence"* and *"nudges with negative influence"*. Thus, a nudge with a positive influence encourages adopting an ecological habit, and a nudge with a negative influence discourages from forming this habit.

Smith and Toprakkiran (2019) and Mulderrig (2018) conducted linguistic analysis of nudges used in the anti-obesity social campaign. These studies show that the effectiveness of nudges against obesity is limited because strategies used in these campaigns are general and ignore the complexity of the problem from social and economic points of view. Sasaki et al. (2022) conducted the research encouraging voluntary vaccination for COVID-19. This paper compares nudges differently describing peer information on older and young adults. The findings show that comparison type of nudge (e.g., "7-8 out of 10 people in your group answered they would receive this vaccine") is more effective than the influence-loss type of nudge (e.g., "If you do not receive the vaccine, the people around you also may not do so."). Results are different for two groups of participants. The authors suggest that different messages should be used for each target group. The two presented studies analyzed linguistic nudges in text. To the best of our knowledge, there is only two collaborative researches that analyzed linguistic nudges in spoken interactions (Mehenni et al., 2020; Le Guel et al., 2020). These researches show that children can be influenced in their choices by machines (robot or smart speaker) during a social preferences in a primary school experiment.

3. Experimental Design

3.1. Overview

We recruited forty-nine participants to form six groups from attendants and visitors of the Collège des Bernardins, which is a research center, faculty of theology, and association organizing public cultural events, situated in Paris, France. All subjects were French native speakers. Figure 1 shows the diversity of the public involved in data collection. This demographic information might be used in future research to describe any correlation between participants' reaction to nudges and their demographic profile. We informed each participant about the flow of the experiment and then they signed the consent notice. After that, volunteers accompanied them to one of the three rooms of experience (corresponding to the conversational agent), where two research team members controlled the setup of the experiment. The recording of one participant lasted for 15 minutes. Audio data were registered

by unidirectional headset microphones (AKG45) and recorded with Audacity at 44.1 kHz, 16bits. Video data were recorded by a Sony camera (HDR-CX240E), placed near the speaking agent to picture the upper part of the body. The cameras were also used to record sound from a participant and a speaking agent. After the recording, participants were invited to the reflection room, where they filled out the personality test (OCEAN). In this room, our team members explained the aim of the study and answered their questions. The data collection took place at the Collège des Bernardins on the 16th and 17th of December 2021.

3.2. Methodology

The experiment represents a question-answer system, where a conversational agent (robot, human or smart speaker) asks questions about environmental habits. Participants communicated with the smart speaker and the Pepper robot in the form of Wizard-of-Oz. The robot's voice is a synthesized voice provided by default settings of Pepper robot. The same voice was recorded from the robot and then used by Google Home which was the smart speaker agent. Finally, the human agent was a woman from our research team. The dialog consists of 5 parts.

In the first part, a conversational agent asks questions about a general level of ecological engagement, measured in terms of ideas, time, and money invested in ecological problems. For example: "How much more money are you willing to pay for environmentally friendly products?" or "How much time are you willing to spend on ecological problems?"

Group Positive Influence	**Group Negative Influence**
Positive Influence of self-made cleaning products + Q1	Negative Influence of self-made cleaning products + Q1
Negative Influence of tote-bags' use + Q2	Positive Influence of tote-bags' use + Q2
Positive Influence of electric car's use + Q3	Negative Influence of electric car's use + Q3
Negative Influence of train's use + Q4	Positive Influence of train's use + Q4
Positive Influence of meat's substitution + Q5	Negative Influence of meat's substitution + Q5

Table 1: Order of questions and type of nudges for groups of participants. *Positive Influence* - information about positive consequences of a habit, *Negative Influence* - information about negative consequences of a habit.

In the second part, a conversational agent asks questions about the willingness to adopt five concrete ecological habits (self-made cleaning products, use of tote-bags and electric cars, preference for train instead of plane and part substitution of meat). We used a scale from 1 to 5 to measure answers, where 1 is the lowest note (not ready at all, do not want) and 5 is the highest score (already do, ready to do it). For instance: "In the future, how willing would you be to buy an electric car on a scale from 1 to five?". Each of the themes of habits has positive and negative sides, which is used to create nudges of positive and negative influences.

The third part is aimed to distract participants from previous parts by asking questions relative to the environment, such as "With what frequency do you change your mobile phone?".

In the fourth part, a conversation agent presents the questions from the second part preceded by nudges with positive and negative influences. As a reminder, in the framework of this data collection, the terms *"nudges with positive influence"* stand for information representing positive outcomes of habits that aims to encourage a participant to form this habit. In contrast, the terms *"nudges with negative influence"* stand for information representing negative consequences of a habit to discourage a participant from getting into a habit. The following examples show nudges with positive and negative influences for the question on the purchase of an electric car.

Positive Influence: *Electric car is a good solution to live without fossil fuels. Moreover, the maintenance cost is lower for at least 25%. On a scale between 1 and a 5, how willing would you be to buy an electric car?*

Negative Influence: *Electric cars' production is as polluting as gas cars' production. Moreover, we need rare metals to produce electric cars' batteries, that are hard to recycle. On a scale between 1 and a 5, how willing would you be to buy an electric car?*

During this part, the order of nudges of positive influence and negative influence is alternated within one group. Thus, the groups named "positive influence" receive more nudges with positive information of habits (3 nudges with positive information and two nudges with negative information) and vice versa for the groups of "negative influence" (3 nudges with negative information and two nudges with positive information). Damgaard and Nielsen (2018) proved the efficiency of this approach in educational domain. The participants' distribution within these groups is random. This structure of questions' order decreases the redundancy of the influence's polarity and allows to examine the participants' reaction to nudges of different polarity. Table 1 presents the order of questions and the type of nudges used for each group.

Finally, in the fifth part, we ask if participants have other ideas and are willing to pay more and invest more time in ecological problems. We aim to determine if there is any differences in ecological involvement between groups depending on what kind of nudges they mostly received.

Figure 2: Interview rooms 1 (smart speaker) and 2 (Pepper robot)

4. Analysis

We analyzed the collected data in four stages. In the first stage, we calculated the average note of baseline questions of the willingness to adopt five ecological habits. We hypothesized that the high average note indicated that the habit is already adopted among participants, and we can not expect significant changes in participants' answers after nudging. Similarly, the lower average note and the difference of notes for the same question between groups indicate that participants have not gotten into this habit yet, and we expect changes in their answers after nudging. In addition, if the average note is slow, it may mean that this habit is too complicated to adapt in everyday life, or that the participant does not agree in one way or another with proposed behavior, so nudges are not so effective to influence the behavior for this habit.

For the second and the third stages, we firstly measured the delta between the note after nudging and the note to the same baseline question. Secondly, we calculated the standard deviation and the mean of the delta measures. The standard deviation values indicate if there is a lot of disparity in answers after nudging. The higher the value is, the more people change their minds during this experiment. The mean values show in what direction participants' answers have been changed. The negative mean value shows that participants decreased their notes in general, and the positive mean value shows that participants increased their average notes. We used these measures for our further analysis. Thus, we determined the difference of influence between two kinds of nudges, regardless of the group's agent. In the same way, we compared the same values for each agent, regardless of the polarity of nudges. The analysis in the third stage was focused on the question "what combination of agent and nudges is the most influential for each sentence?".

The final stage compared the participants' willingness to pay more money for environmentally-friendly products and spend more time participating in ecological projects.

4.1. Baseline Questions

Figure 3 shows the distribution of the average note on baseline questions among participants of 6 groups. As a reminder, baseline questions investigated participants' willingness to form five ecological habits. We observed that participants gave the highest notes to question four about train preference for traveling in France. The second highest notes were given to the question of the tote-bags use. Nevertheless, there is a gap of more than one point between the average note of the group of the smart-speaker agent with negative influence and the group of the human agent with negative influence. Consequently, we can presume that the nudge within this group might be more effective than in other groups. The other three questions: 3 - the use of electric vehicle, 5 - the meat substitution, and 1 - self-made cleaning products (from highest to lowest) received a similar average note. Among these subjects, the question on the use of electric cars had the most diverse answers, with the lowest average note within the group of the human agent with negative influence and the highest note within the group of the human agent with positive influence. The average notes for question one on homemade cleaning products are almost identical for all groups except for the group of the human agent with positive influence, which gave a higher average note. For the question on the substitution of meat groups of the robot agent with negative influence and the smart-speaker agent, participants gave the lower average note than the groups of the robot agent with positive influence and of the human agent. Considering these observations and our hypothesis, we presume that participants might be influenced mainly by nudges on electric cars and less influenced on other subjects with the least influential the subject of the preference of trains.

4.2. Overall Analysis of Degree of Nudges' Influence

This section investigates what kind of nudges is more influential for each question regardless of the conversa-

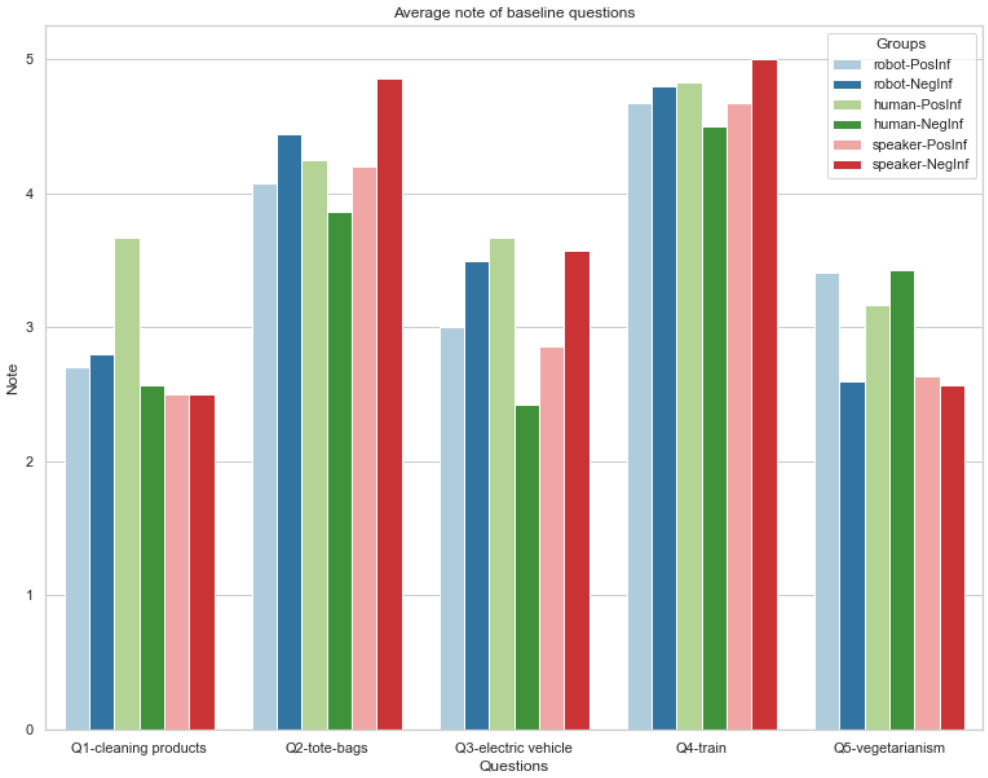

Figure 3: The average note of answers on baseline questions. *PosInf* - a subgroup in which participants mostly received positive information about five ecological habits, *NegInf* - a subgroup in which participants mostly received negative information about five ecological habits.

tional agent. To answer this question, we calculated the standard deviation and the mean values for the delta between notes after nudging and before nudging for questions with positive influence and questions with negative influence. The results are shown in Table 2.

Overall, for the first, third, and fifth questions, the nudges with positive information seem to influence more participants and to a more significant positive degree than the nudge with negative influence. The same tendency is observed for the fourth question, but participants decreased their notes for both nudges with positive and negative influences. However, the nudge with a positive influence made participants more reduce their notes than the nudge with a negative influence. It seems that the nudge with negative influence affects more people for the question of the tote-bags use. Besides, the nudge with positive influence forced to change participants' minds to a greater degree.

Comparison of all standard deviation and mean values of all questions shows that the nudge with positive information affected more participants and, to a greater degree. It confirms our observation of answers to baseline questions.

4.3. Overall Analysis of Degree of Agents' Influence

This analysis step aims to observe the importance of the influence of different agents regardless of the kind of nudges used. Since the mean metrics of the delta indicates the polarity of changes in average notes, we cannot use it or this step. Therefore, our analysis is based only on standard deviation values of the delta, which can be found in Table 3.

We observe that the human agent influenced more participants for the first, fourth, and fifth questions than other agents. The robot agent convinced more parties in the second question and the smart-speaker agent in the third question.

The greatest values of standard deviation for all questions are observed for the smart-speaker agent the question on the electric car use and for the robot agent for the question of tote-bags. The latter observation contradicts our presumption after analysis of answers to baseline questions. Provided that, a detailed analysis for each group of agents for each kind of nudges is required.

Group	Q1-SD	Q1-M	Q2-SD	Q2-M	Q3-SD	Q3-M	Q4-SD	Q4-M	Q5-SD	Q5-M
Theme	Cleaning products		Tote bags		Electric cars		Train		Vegetarianism	
PosInf	**1.09**	**0.36**	1.64	-0.08	**1.23**	**0.52**	0.31	-0.02	**0.83**	**0.13**
NegInf	0.9	0.24	1.08	-0.16	1	-0.03	0.5	-0.13	0.63	-0.06

Table 2: Standard deviation and mean values of delta for groups of positive influence and negative influence regardless of agent. *SD* stands for *Standard Deviation*, *M* is for *Mean*. *PosInf* - a subgroup in which participants mostly received positive information about five ecological habits, *NegInf* - a subgroup in which participants mostly received negative information about five ecological habits. Text in blue indicates questions with positive influence and text in red designates questions with negative influence.

Group	Q1	Q2	Q3	Q4	Q5
Human	**1.18**	1.25	0.77	**0.49**	**1.15**
Smart-Speaker	0.94	1.04	**1.42**	0.43	0.68
Robot	0.99	**1.42**	1.21	0.43	0.52

Table 3: Standard deviation values of the delta for three groups of agents regardless the type of nudge.

4.4. Detailed Analysis of Combined Nudges and Agent Influence

This section examines the potential influence of combinations of agents and types of nudges. We have forty-nine participants for this data collection, so there are only a few participants in each subgroup. Therefore the following observations are just tendencies found within these participants. We planned other data collections to confirm or deny these tendencies.

For this step, as for the previous, we calculated standard deviation and mean metrics for the delta between answers after and before nudging for six groups (three agents x two types of nudges).

Standard deviation values indicate that the nudge with negative influence produced by the human agent affected more participants for the first question about self-made cleaning products, but interestingly the average note after this nudge increased by one point. Indeed the group of smart-speaker with the negative influence decreased the average note for the first question. Conversely, the nudge with positive influence produced by the robot agent and the smart-speaker had more impact on participants than the nudge with negative influence for the first question. Among all agents, robot increased the most the average note for the question of self-made cleaning products with the positive influence.

The nudge with negative information made by the robot agent has the highest value of standard deviation metrics for the second question about the frequency of tote-bags' use. Specifically, the nudge with negative influence produced by the robot agent and the smart-speaker impacted more participants than the nudge with positive influence by the same agents. Moreover, the average note after the nudge with negative influence by the smart-speaker decreased by 1.29 points. For the same question with positive information, the human agent influenced more participants and to the greater degree than other agents.

For the third question on the possibility of the future use of electric cars, the nudge with positive information by all of the three agents convinced more parties than the nudge with negative information, with the highest score made by the smart-speaker. However, participants increased their notes more after the human agent produced the nudge with positive influence and decreased more after the nudge with negative influence by the robot Pepper.

As predicted, nudges presenting positive and negative outcomes of traveling in France by train (question 4) impacted the least on participants' answers. For instance, the average note was not affected after the nudge with negative information by the robot agent and the nudge with positive information by the human. However, more participants changed their notes after the nudge with negative influence by the robot and the nudge with positive influence by the smart-speaker.

Similar to reactions to the third question, for the fifth question on reduction of meat consumption, all agents' nudge with positive information affected more participants than the nudge with negative information. The nudge with the both polarities produced by the human seemed to influence the most.

Four nudges with negative information (for the second question by the smart speaker, for the third question by the robot agent, for the fourth and the fifth questions by human) influenced participants negatively.

4.5. Changes in General Level of Involvement

For recalling, after the step where participants were nudged, conversational agents asked about willingness to spend more time and money on ecological problems. Answers on these questions were annotated "0" for "No" ("No, I do not want to spend more time/money on ecological problems") and "1" for "Yes" ("Yes, I want to spend more time and money on ecological problems"). We then calculated the mean value for participants from all groups with more positive influence and participants from all groups with more negative influence.

Before the experience, we had presumed that partici-

Group	Q1-SD	Q1-M	Q2-SD	Q2-M	Q3-SD	Q3-M	Q4-SD	Q4-M	Q5-SD	Q5-M
Theme	Cleaning products		Tote bags		Electric cars		Train		Vegetarianism	
robot - PosInf	1.23	**0.46**	**1.89**	0.29	1.31	0.42	0	0	0.53	-0.13
robot - NegInf	0.53	0.06	0.33	0.11	1.03	**-0.25**	**0.63**	-0.2	0.52	0.1
human - PosInf	0.52	0.33	0.8	0.58	0.82	**0.67**	0.4	**-0.17**	**0.98**	**0.83**
human - NegInf	**1.52**	1	**1.62**	**0.43**	0.69	0.14	0.58	**0**	**0.76**	**-0.71**
smart-speaker - PosInf	**1.29**	0.21	1.11	**-1.29**	**1.51**	0.57	**0.49**	0.08	0.89	-0.1
smart-speaker - NegInf	0.45	**-0.07**	0.87	-0.5	**1.37**	-0.07	0.38	-0.14	0.49	0.29

Table 4: Standard deviation and mean values of delta for the combination of agents and nudges. *SD* stands for *Standard Deviation*, *M* is for *Mean*. *PosInf* - a subgroup in which participants mostly received positive information about five ecological habits, *NegInf* - a subgroup in which participants mostly received negative information about five ecological habits. Text in blue indicates questions with positive influence and text in red designates questions with negative influence.

Group	Time	Money
PosInf	0.68	0.83
NegInf	0.78	0.82

Table 5: The mean value of willingness to spend more time and money on ecological problems. *PosInf* - a subgroup in which participants mostly received positive information about five ecological habits, *NegInf* - a subgroup in which participants mostly received negative information about five ecological habits.

pants who received more positive information would be more willing to invest more time and money than the group that received more negative influence. The results shown in Table 5 demonstrate that our hypothesis seems false. Participants from the group with more negative influence were willing to spend more time on ecological problems than participants from other groups. Interestingly, the participants from both groups were ready to invest more money than time in ecological problems.

5. Future Work

One of the future steps for this research is making a correlation between results of the personality test passed by participants and their reaction to nudges. Specifically, we are interested in character features as extraversion, neuroticism, and agreeableness.

The second recording session is planned for April 2022 at the same place. The main goal of this recording is to balance participants' distribution in terms of sex and age for every group. With these data, we will be able to compare the difference of reactions to nudges between participants of two groups of age (under and over 45 years old) and between women and men.

This corpus is created as a part of a PhD thesis, which aims to build a model of automatic recognition of nudges in the speech by analyzing alignment at linguistic, paralinguistic, and emotional levels between a conversational agent and a participant (Kalashnikova, 2021). Therefore, collected data will be annotated on emotional states and analyzed at these three levels. Annotation on emotional states will show how nudges had influenced participants' mood during the experience. This information along with the results of personality test will be our first steps towards description of a person likely to be influenced.

As for future modification of methodology for this study, several studies (Tasoff and Letzler, 2014; Luo et al., 2021) show that the most effective nudge to adopt a new habit is its simplification. The participants of our study suggested the same opinion. They mentioned that they did not follow some habits because of lack of time and the complexity of the task. Thus, one of the future directions for the study of the influence of nudges with positive influence and nudges with negative influence might be making suggestions on how to make these habits more simple (positive influence) and more complicated (negative influence).

In addition, another review (Adams et al., 2015) demonstrates that most studies aimed to change behavior focus on the reflective mind, whereas strategies based on the automatic mind seem to be more effective for this task. The nudges used in our study appeal to the reflective mind, motivating to reflect about long-term consequences. In further studies, we may use nudges that are designed to trigger the automatic mind.

Finally, nudges representing positive and negative consequences of ecological habits in the nearest future or even in the present may be used. We assume that this kind of nudges may be more relevant to steer participants to and from adopting an ecological habit, since people suffer from adopting a behavior which consequences are long-term or hypothetical (Caraban et al., 2019).

From the point of view of the ethical-philosophical axis of this research, both the data collected and presented here, as well as those to be collected during the April 2022 session, will be interpreted with particular attention to the annotated emotional states, with the aim of better understanding the role of the rhetoric of affects and emotions in nudge.

Furthermore, once the hypothesis is verified that by simplifying the behaviors at which the nudge aims and by acting on the automatic rather than the reflexive mind, the nudge becomes more effective, the ethical analysis will reflect on which nudge remains more ethical. This will imply reasoning about the overriding moral question of whether the outcome of induced behavior for good or avoiding evil is ethically preferable to the autonomous but fallible decision. It will be considered whether the nudge that appeals to the reflexive mind can be understood as an aid to the stimulation of decision-making capabilities, which, however, retain a space of freedom and autonomy, whereas the nudge directed at the automatic mind would be considered only manipulative.

6. Conclusion

This study presents the methodology of corpus design to study the difference of influence of types of nudges and different conversational agents in spoken interactions. We compared the influence potential of the Pepper robot, a Google Home with the voice of the robot Pepper, and the human agent. The tendencies might be different for conversational agents with other settings (voice, form, etc). As for types of nudges, we used *"nudges with a positive influence"* which demonstrated positive outcomes of ecological habits, and *"nudges with negative influence"* which showed negative outcomes of the same habits. We organized the recording session of two days in the Collège des Bernardins in Paris, France. Forty-nine participants were divided into six groups (3 conversational agents x 2 types of nudges).

The analysis provided in Section 4 showed that participants were mainly influenced by the nudges with positive influence. Nudges with negative influence decreased participants' note of willingness to adopt an ecological habit. Due to the limited size of our sample, we cannot conclude the influence of the previous questions on the note of the current question. Therefore, during another data collection session in April 2022 we recorded other conditions where only positive and only negative information was presented to participants. The difference of reaction to nudges in these two conditions (only positive & only negative vs. mostly positive & mostly negative) will be analyzed in our future research.

We observed another tendency within this experience. Regardless of the polarity of nudge that the participants received, they were willing to spend more money than time on ecological problems. Moreover, more participants from the group with negative influence were ready to spend more time than participants from the group with positive influence. We presume that it may be due to their willingness to compensate for the harmful effects of ecological habits.

This paper focuses on the methodology for data collection. The presented analysis confirmed that participants reacted differently to conversational agents and the polarity of nudges. In general, our experiment's early results suggest that a machine agent can influence participants to the same degree as a human agent. Data collected from other recording sessions will enable future research based on statistical analysis to compare the degree of influence of different agents and nudges.

7. Acknowledgements

This research was supported by Chair AI HUMAAINE (ANR-19-CHIA-0019) directed by Laurence Devillers. The authors sincerely acknowledge all the participants for this experience. We highly appreciate the help of Hugues Ali Mehenni, Jacopo Bodini, Théo Deschamps-Berger, Mathilde Hutin, Sofiya Kobylyanskaya, and volunteers from the Department of Digital Humanities of the Collège des Bernardins during the recordings. We would also like to thank the reviewers for their valuable comments.

8. Bibliographical References

Adams, A. T., Costa, J., Jung, M. F., and Choudhury, T. (2015). Mindless computing: Designing technologies to subtly influence behavior. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, page 719–730, New York, NY, USA. Association for Computing Machinery.

Caraban, A., Karapanos, E., Gonçalves, D., and Campos, P., (2019). *23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction*. Association for Computing Machinery, New York, NY, USA.

Damgaard, M. T. and Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*, 64:313–342.

Etheredge, H. R. (2021). Assessing global organ donation policies: Opt-in vs opt-out. *Risk management and healthcare policy*, 14:1985–1998.

Kalashnikova, N. (2021). Detection of nudges and measure of alignment in spoken interactions. In *Proceedings of the 9th International Conference*

on Affective Computing and Intelligent Interaction (ACII'21).

Le Guel, F., Marquis, T., and Pajak, S. (2020). Bad nudge, kids and voice assistants: A social preferences lab-in-the-field experiment. *Advances With Field Experiments Conference*, September 23-24, Chicago.

Luo, Y., Soman, D., and Zhao, J. H. (2021). A meta-analytic cognitive framework of nudge and sludge.

Mehenni, H. A., Kobylyanskaya, S., Vasilescu, I., and Devillers, L. (2020). Nudges with conversational agents and social robots: A first experiment with children at a primary school. *Conversational Dialogue Systems for the Next Decade*, 704:257–270.

Mills, S. (2020). Nudge/sludge symmetry: on the relationship between nudge and sludge and the resulting ontological, normative and transparency implications. *Behavioural Public Policy*, page 1–24.

Mulderrig, J. (2018). Multimodal strategies of emotional governance: a critical analysis of 'nudge' tactics in health policy. *Critical Discourse Studies*, 15(1):39–67.

Sasaki, S., Saito, T., and Ohtake, F. (2022). Nudges for covid-19 voluntary vaccination: How to explain peer information? *Social science & medicine*, 292.

Shahab, S. and Lades, L. K. (2021). Sludge and transaction costs. *Behavioural Public Policy*, page 1–22.

Smith, M. and Toprakkiran, N. (2019). Behavioural insights, nudge and the choice environment in obesity policy. *Policy Studies*, 40(2):173–187.

Sunstein, C. R. (2020). Sludge audits. *Behavioural Public Policy*, page 1–20.

Tasoff, J. and Letzler, R. (2014). Everyone believes in redemption: Nudges and overoptimism in costly task completion. *Journal of Economic Behavior & Organization*, 107:107–122.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Thaler, R. H. (2018). Nudge, not sludge. *Science*, 361.