Contrasting Sentences as Engagement Strategy

Luca Garello, Francesco Grella, Stefano Castagnetta
DIBRIS - Department of Computer Science, Bioengineering, Robotics and System Engineering
University of Genoa, Italy - A.Y. 2018-2019
lucagarello@outlook.com
francesco95grella@gmail.com
stefanocastagnetta@gmail.com

***Abstract*—The concept of conversation is an abstraction ingrained in human mind which resumes a natural verbal interaction between one or more other individuals: the extension of such a natural ability to artificial beings just as robots is clearly not trivial. The goal of this project is to modify the conversational capabilities of Pepper Robots, already involved in the Caresses Project voted to study and improve verbal interactions with elder people, and make them suitable for having a flowing and interesting speech with younger people. In particular through modifications on the ontology which handles the dialogues of the robot, the intent is to improve Pepper's personality in a contrasting way with respect to its human interlocutor, for instance by adding phrases induced by a positive response of the user to a question which are in contrast with his opinions. This new contrasting behavior of the robot has been tested 'on the field' with subjects with an average age of 23 years old, and the results have been confronted with the ones obtained with a control group, which has interacted with the original ontology.**

I. INTRODUCTION

For what concerns the context of verbal interactions, there are several psychological considerations that can be done; in particular there are many individual factors which can drive a conversation and influence the behavior and the reactions of the participants.
For instance on a cultural point of view there could be differences in reacting to some specific topics: the interest, indifference or even inadequacy with respect to a certain argument could vary dependently on participants own culture.
Another influencing factor which could drive verbal interaction in a direction rather than another one is certainly the age of the people involved: it's easy to notice, even in everyday life, that elder people have a different way of communication with respect to middle aged and young people in terms of topics, tones and obviously politeness.
There are many rules that have to be respected when interacting with people belonging to different age ranges, which could be very strong for some cultures (Japan) or weaker (Europe) but always well fixed in our behavior.
To get to the point of the project, let's consider a robot designed and programmed for verbal and non-verbal interaction with humans: in particular the starting point of our work was a culturally competent robot with a knowledge base conceived for interaction with elder people [2].

Firstly it must be noticed that importance has been given to the politeness of the sentences, according to what said before; moreover since the robot it's designed for interacting with elders, interactions are mostly driven on age-specific topics, with particular accent on the assistance-related topics like health status or medicine remind.
It's obvious that a robot which performs such interactions must have a 'smooth personality' and be compliant with the user, and must be able to acknowledge when the conversation is becoming annoying and distinguish when the human wants to be left alone to rest or rather it's feeling alone and simply wants an entity which listens to him/her.

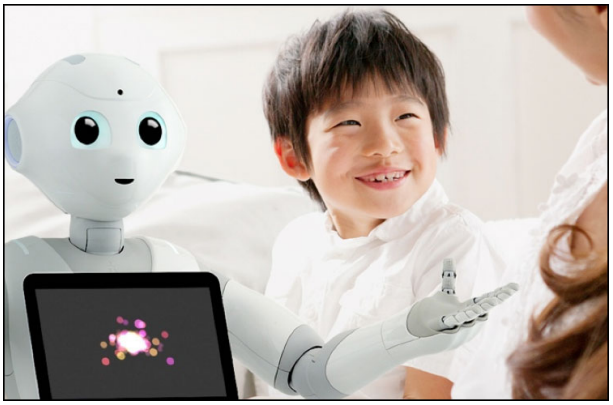

Fig. 1: Pepper Robot

Anyway the previous considerations are not valid anymore when the age of the target is drastically reduced: assuming that we are neglecting interactions with middle aged people because such a study is way far from the scope of this project, we thought that a totally different behavior is expected when the robot has to interact with young people, intended as people which are at most 27/28 years old.
In this case the conversation topics must be completely different and the interaction itself is expected to be way more dynamic and interesting.
To clarify this conceptualization let's consider a 20 years old guy which has its own opinions on topics like music, food or sports and let's imagine that the robot follows a dialogue pattern which is the same used when interacting with elder people: even assuming that the topics of the conversation are different and closer to the guy's preferences, the tone of the conversation and the manners shown by the robot will soon

result inadequate to the context of that particular interaction, and the user clearly will be bored after a short dialogue session, because the high politeness of the sentences imposes the appeasement.
What we have tried to prove, instead, is the fact that the conversation could become more natural, interesting and even entertaining by adding proper sentences to the dialogue algorithm which are thought to be in contrast with the positive or negative answers given by the user to robot's questions about various topics.

II. STRUCTURE OF A CONVERSATION

When a robot talks with a human, it is helpful to know what type of conversation will be engaged: a social Robot should be able to adopt different types of conversations according to the context, the mood and the personality of the human interlocutor. In scientific literature there are many research that try to classify the type of a conversation.
One interesting exploration of the structure of casual conversations is the one of David W. Angel [1].

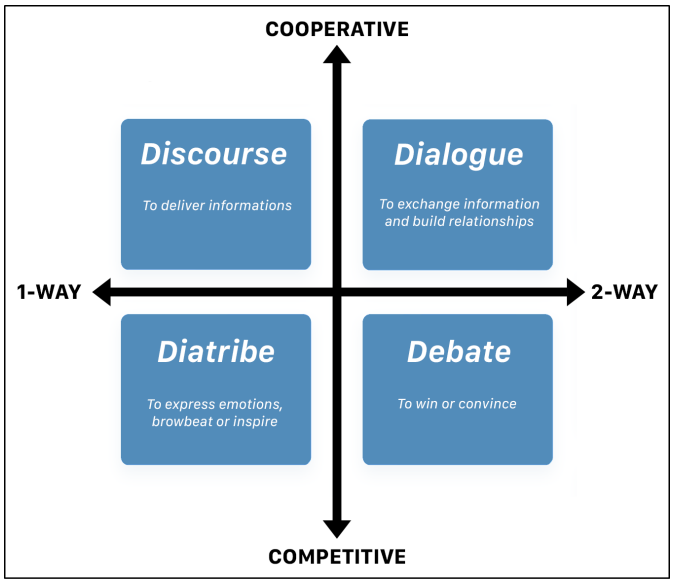

Fig. 2: The Four Types of Conversations

According to his research the first important distinctions is the conversations direction of communication, that can be one-way or two-way street.
In a two-way conversation, participants are both listening and talking while in a competitive conversation, people are more concerned about their own perspective, whereas in a cooperative conversation participants are interested in the perspective of everyone involved.
The second distinction is the tone/purpose of the conversation that can be *competitive* or *cooperative*. Basing on these directions and tone four possible types of conversation can defined.

- *Debate* is a competitive, two-way conversation. The goal is to win an argument or convince someone, such as the other participant or third-party observers.
- *Dialogue* is a cooperative, two-way conversation. The goal is for participants to exchange information and build relationships with one another.
- *Discourse* is a cooperative, one-way conversation. The goal to deliver information from the speaker/writer to the listeners/readers.
- *Diatribe* is a competitive, one-way conversation. The goal is to express emotions, browbeat those that disagree with you, and/or inspires those that share the same perspective.

Under the assumption that assistive robots should promote 2-way conversations we are interested in the development of engaging dialogues and debates (right side of the diagram).

III. CONTRASTING SENTENCES

The framework that ensures the verbal capability of the robot relies on a three-layer ontology for storing concepts of relevance, culture specific information, person specific information and preferences. The sentences stored in the data properties hasQuestion, hasPositiveSentence, and hasNegativeSentence ensure that the robot can discuss the instance they refer to with the user.

Data property hasQuestion contains the question(s) the robot should use to ask the user about any instance subsumed by Topic , while the data properties hasPositiveSentence and hasNegativeSentence contain sentences that the robot can use to express, respectively, a positive or a negative attitude towards the instance subsumed by Topic.

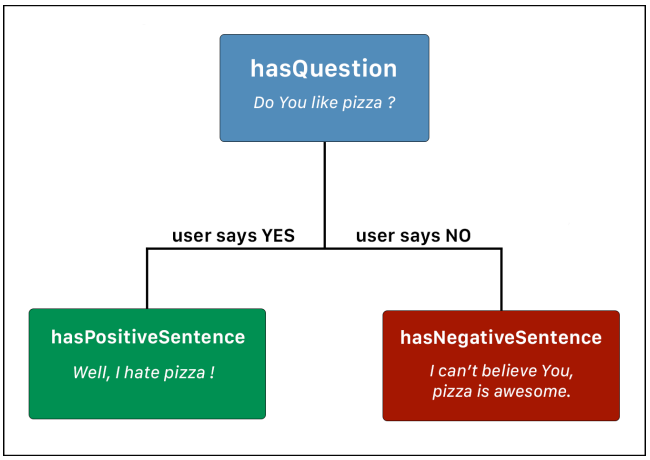

Fig. 3: Simplified scheme of the contrasting conversation implemented

Fig. 3. shows a simplified scheme of Robot Question/answers architecture. Based on the user's answer the Robot is able to talk back with a contrasting sentence.

For our research we have introduced two kind of contrasting sentences:

- *Contrasting Negative sentences:* when Robot disagrees on a negative opinion of the user.
- *Contrasting Positive sentences:* when Robot disagrees on a positive opinion of the user.

There isn't a sharp rule that defines "contrasting sentences" or classifies them, for this reason we have tried to group them into four categories.

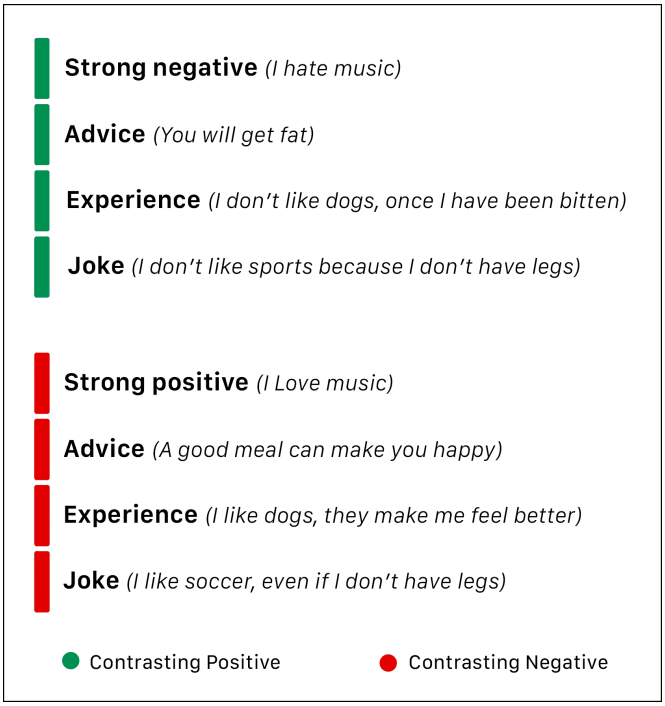

Fig. 4: Examples contrasting Positive and contrasting Negative Sentences

As Fig. 4. shows, contrasting sentences (positive or negative) can be in turn presented by the robot in different ways:

- *Strong negative/positive:* Contrasting sentence with strong temperament.
- *Advice:* Advice that warns the user contrasting his answer.
- *Negative experience:* Answer that is justified by an experience or a more general reasoning.
- *Joke:* Answer softened by a joke.

In some cases the categories proposed can overlap but we think that this representation can help developers to define different competitive robot personalities.

IV. EXPERIMENT SETUP

A. Research Questions

- People find conversations to be more entertaining if the robot disagrees with them from time to time.
- People find conversations to be more entertaining when carried with a humanoid rather than with a vocal assistant.

conversations with Pepper are more entertaining than with other devices (mainly vocal assistants)

B. Hypothesis

We want to address the way the entertainment of a conversation (dependent variable) is related to the attitude of the robot (independent variable). Since we focus only on different attitudes of the robot, we don't expect to have any confound.

C. Conditions

We assess our questions and hypothesis by comparing two scenarios which differ for the attitude of the robot towards the user: in the first scenario the human interacts with the default ontology of the CARESSES project, which is rather compliant with the user. In the second scenario the robot is not likely to agree with the human as presented in Section II.

D. Experimental Units

A total of 15 individuals completed the interaction with the Robot. There were 10 females and 5 males; Respondents ranged in age from 20 to 24 (n = 11, 73.3%) and 25 to29 (n = 4, 26.7%). Half of the participant had ever interacted with a physical robot (50.0%) and from the questionnaires it appeared that the majority of them uses smart assistant (e.g google assistant, alexa, siri) to get information rather than to have a conversation. Almost all participants were university students (93%).

E. Study Design

We assigned the two conditions between-subjects (i.e. each unit is assigned one condition) in order to not introduce learning confounds related to the order in which the conditions are presented to the unit.

Although we know that this procedure usually requires an higher number of units to obtain statistical relevance, we want to point out that this is intended to be a pilot study and validation could be obtained through future work.

F. Variables

In addition to the variables introduced in IV-B we also have some covariates. While some of them were hard to control (e.g. frequency of interactions with robots, previous experience with vocal assistant), we tried to keep the age and gender as homogeneous as possible between the two groups.

G. Measures and Metrics

The interaction with the Robot lasted 10 minutes at the end of which each participant completed a questionnaire (refer to Tab. I).

The evaluation form has been designed for collecting basic data on the subjects and several specific data about the interaction with Pepper.

In particular the subject is expected to express a preference in a linear scale of ten values regarding some outcomes of the interaction, focusing on individual impressions.

TABLE I: Statements listed on subject questionnaire

Subject questionnaire
(1) Feeling of being understood
(2) Level of interest of the conversation
(3) Level of coherency of the answers
(4) How comfortable was the conversation?
(5) Level of emotional involvement
(6) Level of enjoyment of the conversation
(7) How natural was the conversation?
(8) Level of courtesy

[3] showed evidence of relation between time perception and entertainment. In addition to the questionnaire, we asked the subjects to give an estimate of the duration of the interaction. This measure is independent with respect to the answers of the questionnaire and in someway it is more objective.

V. RESULTS

Although we kept the two groups as homogeneous as possible in terms of controllable covariates (i.e. age and gender), given the limited population we couldn't impose strict prerequisites of participation. As a result, while in the experimental group every participant has had previous experiences with vocal assistants, in the control group subject S-7 never had the chance to verbally interact with a device. For this reason in the next 2 subsections, other than the groups results, we will report her individual results to investigate the possibility that her answers were influenced by this being her first vocal interaction.

An insight at the frequency of interaction of the population is displayed in Fig. 5. The picture shows that almost half of the participants had never interacted with a physical robot before and only one of them does it regularly.

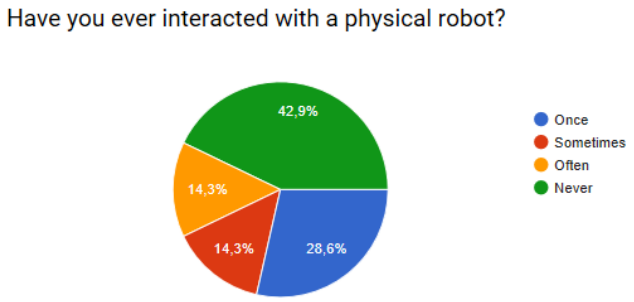

Fig. 5: Frequency of interactions with a physical robot

A. *Responses to Questionnaires*

In Fig. 6 we display the responses to the questionnaire about the interaction with Pepper, comparing the answers of S-7 with the mean of those of the rest of the control group. The level of agreement of S-7 resulted much higher than that of any other subject. Given that she was the only one to have never interacted with a vocal assistant, in the following we don't include her in the mean of the control group. Moreover, clearly she did not provide responses for the assessment of previous interactions.

Note: vertical error bars indicate standard deviation.

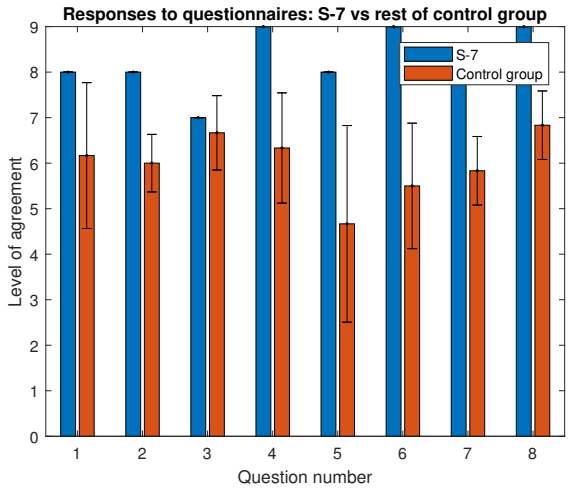

Fig. 6: Responses to the questionnaire about the interaction with Pepper of the control group

Given that we want to find out whether people find a conversation more entreating when the robot disagrees with them or not, we compared the responses of the two groups to the questionnaire about the pepper robot. From the graph showed in Fig. 7 it's possible to see that there isn't a relevant difference between the answers of the two groups.

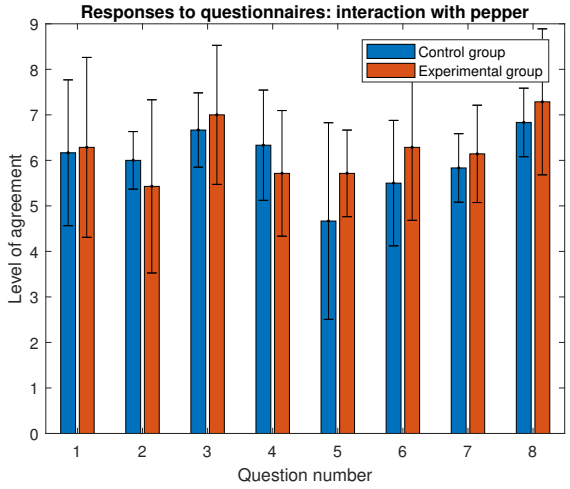

Fig. 7: Responses to the questionnaire about the interaction with Pepper of the control group and the experimental group

Moreover we wanted to compare previous interactions to the one with Pepper. In Fig. 8 we show the mean of the responses of the whole population on these two topics. From the graph

it's possible to see that on average the subjects agreed more with enjoyment statements related to the interaction with Pepper with respect to the previous ones.

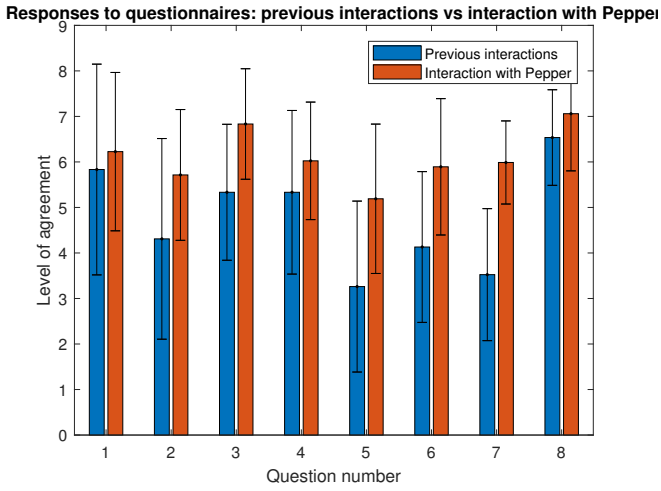

Fig. 8: Responses to the questionnaire about previous interactions and the one with Pepper

Finally, in Fig. 9 we report the comparison of the mean responses on previous interactions between the two groups. The graph shows that subjects from the experimental group on average agreed less with enjoyment statement about previous interactions.

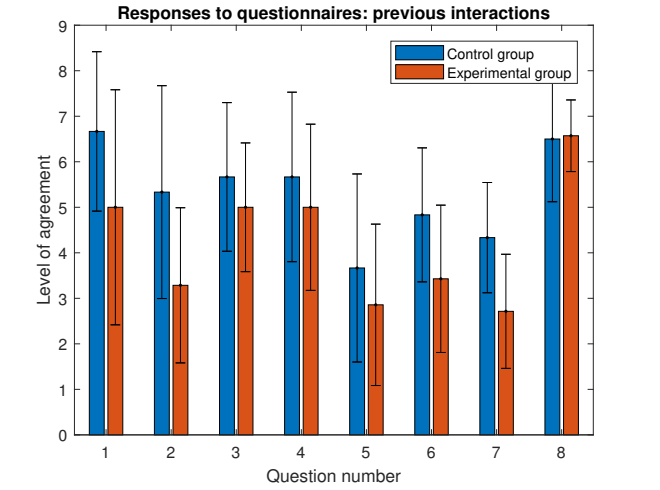

Fig. 9: Responses to the questionnaire about previous interactions of the control group and the experimental group

B. Time Perceived

The mean time perceived by the control group was 10.3 minutes (std = 2.58 minutes), while for the experimental group was 9.0 minutes (std = 3,27 minutes). The perceived time of subject S-7 (not included in the control group mean and std computation) was of 3 minutes.

VI. DISCUSSION

The population tested in this pilot study is fairly limited. Therefore the goal of this discussion on the results is not to draw conclusions on the subject but rather to make considerations and propose metrics to introduce in a future scaled experiment.

A. Responses to Questionnaires

The research question that we posed was that a robot with a "contrarian" personality would lead more entertaining conversations. From the responses to the questionnaire (Fig. 7) we didn't find a neat difference between the appreciation of the two behaviors. In the following we propose some possible response bias that could have had a levelling-out effect.
As a first consideration, we suppose that 10 minutes of conversations may not be enough to trigger a different perception of enjoyment of the conversation. Indeed a subject which never interacted before with a humanoid is still excited that he's able to speak with a robot and therefore doesn't get bored. Additionally we realized that even if all our subjects had an English certification, the language could have been a barrier, which led to some of the jokes not being caught. Finally as it often happens, there is a chance that the subjects of the control group still gave positive evaluations to the "follower" personality as they thought we would be pleased by that (social desirability bias).
For future experiments we suggest to increase the time of the conversation, to keep track of the number of times the user asks: "Could you repeat?", and to have a balanced number of positively and negatively worded questions as well as to isolate the subject when is taking the survey to mitigate the effect of the social desirability bias.

An additional remark on the results of the comparisons between the two personalities is that the mean answers to question 8: "Level of courtesy" were comparable. This is a good hint to the fact that if the robot disagrees with the user more often, it doesn't necessarily mean that it is perceived as rude.

When we compared the answers about previous interactions between the control and the experimental group (Fig. 9), we found that people who interacted with the "contrarian" personality gave worse scores on average for every question. This result was at first unexpected because previous experiences are clearly not related to the different experiences they went through during the experiment. A possible answer to this could be that the interaction with the "contrarian" pepper was more entertaining, thus influencing negatively their judgment of previous interactions and shifting down their level of agreement in general. In any case, this behaviour should be investigated in a broader study.

The second question that we wanted to prove was that conversations with Pepper are more entertaining than with other devices (mainly vocal assistants). From the results (Fig. 8) it seems that this tendency is confirmed, in particular for question 7: "How natural was the conversation".

From the comparison of the responses of Subject-7 with respect to the ones of the rest of the control group (Fig. 6), we can guess that the fact that she had never verbally interacted with a devise created an amazement effect which boosted her level of agreement.

B. Time Perceived

The results show a slight shift of perception of time for what concerns subjects of the experimental group (1 minutes). On the other hand subjects in the control group seemed to be able to correctly guess the duration of the interaction. While it is difficult to determine how great the shift in perception should be to be considered relevant, we think that the results obtained should lead to future experiments to evaluate more in depth this phenomena.

In addition we note that subject S-7 gave responses in the questionnaire that showed a high level of entertainment and she evaluated the time spent with the robot to be 70% less than what it actually was. This could be an indication of the relevance of this metric.

For future studies we propose to have a real duration of the experiment different then 10 minutes as it is common to approximate to the nearest 5 or 10. This could have had a centering around 10 effect on our results.

VII. Conclusions

At present, the trend is to give a ”follower” personality to devices with natural language capabilities. In this work we proposed an ontology which contains contrasting answers, leading to a ”contrarian” personality. Moreover we designed a pilot study to test if said personality leads more entertaining conversations.

Although some of the results were promising, a study with more subjects should follow to provide statistical relevance, and an extension of the set of topics for which contrasting answers are given to the interlocutor could allow the exploration of a deeper and more heterogeneous dialogue, with the objective of reaching a natural conversation with an informal tone.

We retain that the idea developed along the project could be the basis for several different studies related to human-robot interaction, like the possibility of experimenting mixed initiative dialogues, having the robot which expresses its preferences before proposing any topic to the human.

Finally we think that an ”independent personality”, where the robot sometimes agrees and sometimes doesn’t, should also be tested and compared with the other two personalities presented in this work.

References

[1] David W. Angel. The four types of conversations: Debate, dialogue, discourse, and diatribe. Website, 2016.

[2] Carmine Recchiuto Edouard Lagrue Barbara Bruno, Roberto Menicatti. Culturally-competent human-robot verbal interaction. In *2018 15th International Conference on Ubiquitous Robots (UR)*, 2018.

[3] A. M. Sackett, M. Meyvis, L. D. Nelson, B. A. Converse, and A. L. Sackett. Youre having fun when time flies: The hedonic consequences of subjective time progression. *Association for Psychological Science*, 2010.