On effective prototype sessions for the e-book of the future

Cindy Berghuizen, Chiel Peters, Mary Gouseti,
Omar Pakker, Fabiën Tesselaar

December 14, 2013

1 Introduction

Prototyping is a vital eliciation technique for Requirements Engineering. In this document the (un)effectiveness of this technique is researched. When does prototyping provide good results? Two seperate experiments on prototyping were performed in order to assess particulair situations.

The second section contains possible cases that pose a threat to prototyping. The third sections contains the cases which aid in the effectiveness of prototyping. The first expirement using anchoring and group dynamics is given in section 4 and the second follow-up expirement on group dynamics is stated in section 5.

2 Theory

In this section multiple subjects are stated that will likely negatively impact a prototype session. These subjects are based on the theory from Kahneman's book, *Thinking fast and slow*.

2.1 Bad Target Group

The participants in the prototyping session are of vital importance. Their backgrounds will determine the kind of feedback that will come out of the session. Two categories of participants will most likely not give useful feedback. These are:

- Insiders. Participants that already have a stake in the product
- Fanboys/girls. Participants that already like the product, because of the company or any other particular reason.

The insiders are bad participants because they are biased. They can only observe the inside view. Kahneman states the exact problem on the inside view:

He's taking an inside view. He should forget about his own case and look for what happened in other cases.

In a prototype session all the cases of the inside view are probably already covered by the prototype; however the outside view will render cases that give useful feedback. Having only insiders in the prototype session will thus hurt your feedback results.
The second group of participants is fanboys/fangirls. These participants will most likely fall for a different kind of fallacy, namely answering the easier question. Instead of looking at the product from an unbiased perspective they will already like product. This is because, they replace the question *How much do I like this prototype?* with an easier question: *How much do I like the company or their earlier products?*. This based perspective will negatively influence their ability to find new cases or flaws in the prototype.

2.2 Cognitive Ease & Affect

"When you are in a state of cognitive ease, you are probably in a good mood, like what you see, believe what you hear, trust your intuitions, and feel that the current situation is comfortably familiar."
"The affect heuristic is an instance of substitution, in which the answer to an easy question (How do I feel about it?) serves as an answer to a much harder question (What do I think about it?)."
Setting an environment in which the subjects experience cognitive ease may lead them to be more receptive of the prototype. To accomplish that, the environment of the experiment will be comfortable and the participants may be offered food or listen to music or even asked to smile throughout the experiment. According to the first quote of this section, they will not be critical towards the prototype, they will accept what is suggested by the experiment and the prototype will seem familiar. Moreover, they will feel more comfortable using it than they would have. Finally, if they have to answer the questionnaire in the end of the experiment, they will have associated the memory of using the prototype with their good mood so they will be even more positive towards the prototype assuming they use the affect heuristic.

2.3 Deplete System 2

Ego depletion: *"if you have to force yourself to do something, you are less willing or less able to exert self-control when the next challenge comes around"* .
Ego depletion might be a way to let the participants easier accept the prototype as a good one. First let the participants do something that takes a lot of effort, watch a movie, try not to focus on something distracting etc. Then the prototype will be shown. If the only parts of the system shown are the parts that are good

(as explained in 1.6) the participants will easier accept the system to be good overall and do not criticize it.

2.4 Anchoring

A bad prototype session would be to bias participants to like the prototype. One way this can be achieved is the *Anchoring as priming effect*. Before starting the prototype session the participants are shown fake results from a hypothetical prior prototype session. These results show that a separate group really like the prototype and on average gave very positive feedback on it. Although the new participants should not be affected by these results, the anchoring effect suggests that they are more likely to give more positive feedback as well. In the worst case the new participants will exactly match the fake results resulting in a completely useless prototype session.

2.5 Availability

"The availability heuristic, like other heuristics of judgment, substitutes one question for another: you wish to estimate the size of a category or the frequency of an event, but you report an impression of the ease with which instances come to mind."
Availability can be used to influence the results of the prototype session in a way that is desirable. For instance, the answer of a participant to the question whether she would use the prototype in her everyday life can be manipulated; if different situations that the participant could use an eTextbook, such as while traveling by train, or when moving to a summer house, are mentioned before the question, the answer would be more positive. Similarly, if situations that she could not use an eTextbook are mentioned before the question, the answer would be more negative. This would happen because the references before the question change the ease that examples of using eTextbooks in everyday life come to the mind of the participant.

2.6 Priming / Framing on the good features / WYSIATI

Based on the WYSIATI principle, if we only show the good parts, the features that work well and are easy to use, the participants will likely give more positive feedback. The parts of the prototype that do not work well or are forgotten will be completely ignored. Therefore the participants will not notice the faults and limits of the system and will believe that the software works perfectly.

3 Good prototyping

We believe a good prototype session has the following:

- Prototype session closely related to reality

- Intervene / let the participants experiment
- Priming creativity

3.1 Outcome

A good prototype session will provide you with useful information. The participants will experiment with the software / object, in the best case integrate it in their daily life for a little while. While experimenting and using the device they will find out what works well, what does not work, what they miss and what parts are not used at all. This way they can provide better feedback and critique. Also, a successful prototype will uncover the non-functional requirements. For example: the participants may like the system but you find out that they will not buy it. Now work can be done to find out why people won't buy the system so you improve these factors and make sure the system will be a success.

4 Anchoring Experiment

4.1 Setup

A bad prototype session would be to bias participants to like the prototype. One way this can be achieved is the *Anchoring as priming effect*. Before starting the prototype session the participants are shown fake results from a hypothetical prior prototype session. These results show that a separate group really (dis)like the prototype and on average gave very positive(negative) feedback on it. Although the new participants should not be affected by these results, the anchoring effect suggests that they are more likely to give more positive(negative) feedback as well. In the worst case the new participants will exactly match the fake results resulting in a completely useless prototype session.

This is the experiment we conducted to test the anchoring theory. As described in sections 2.4 and above, the participants are required to fill out a questionnaire in which they rate the presented features on a scale 1 to 10. This survey starts with a few lead-in questions and then moves on to the features, after the features there is a open text question for feedback and comments and the final question asks the participants to give their overall rating for the eTextbook. The questions display the averages of a 'hypothetical' prior experiment.

To measure the anchoring effect we created two questionnaires, both using the same questions. While the questions are the same, the averages for the questions are swapped; question 'x' is rated high in one and low in the other. This setup can be seen in Figure 1. After the experiment we measure the average and median of the answers for the high and the low one to see if the anchoring had any effect on our participants.
Our results show that the anchoring effect is obvious in many cases. Only one question has an inconclusive difference and the deviation cannot be contributed

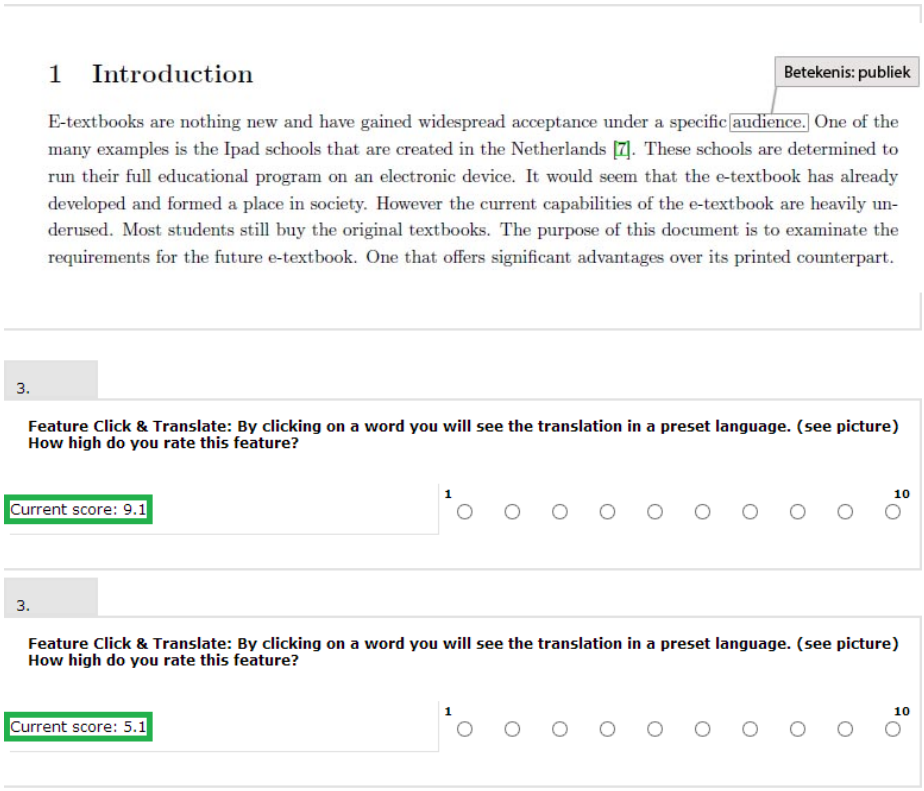

Figure 1: The same question, different anchoring point per questionnaire.

to the anchoring effect for sure.

4.2 Results

This experiment was intended to be an anchoring experiment. However, by checking the results we also saw other theories from Kahneman book put in practice. The results presented in the following sections correspond to 40 participants and the answers were collected from 4/12/2013 to 5/12/2013. The survey is still online and we continue gathering data.

4.2.1 Anchoring

The way to introduce the anchors in out survey was a challenge. Our goal was to present it without drawing a lot of attention on it, so a participant would not reject it. We introduced the survey by explaining that its goal is to test the results of a previous prototype session and we provided the average for every feature. We decided to show the current score on the side of each question because people are familiar with ratings since they are used a lot on the internet. The choice of the anchors was tricky because we wanted them to be believable according to our intuition but we also wanted to preserve a distance between the anchors to get more noticeable results. We decided to approach this separately for each feature and play with different distances.
Although the features were the target questions of our anchoring experiments, in order to be consistent we used the anchors in the general questions too. In this case, we were more careful to provide more believable anchors in order to establish the trust of the participant to the anchors. In Figures 2 and 3 we can see the results of each survey and how the anchoring affects the results.

Anchoring Index. The anchoring effect is measured in percentages by dividing the differences between averages by the difference between the anchors and is called anchoring index. For example, the difference between the anchors 3 and 9 is 6 and the difference between the averages 5 and 7 is 2. Then, the anchoring effect is 2/6, or 30%. An anchoring effect of 100% means the subjects adapts the anchor point and an anchoring effect of 0% means the anchor has no effect on people.
In Figure 4 we see that there was almost no anchoring affect on the question about the dictionary feature (only 3.5%). This is probably because students already know the feature and like it. We will further discuss that in the following section.
The question about the interactive figures had a very strong anchoring effect, it was 53%. This is probably one of the examples that people in general do not have a high opinion of so they are easily influenced by the anchor.
Overall we can conclude the anchors did effect the answers to the questions in almost all of the questions. Therefore, no anchors should be put when doing an prototype session. If people do not have a very strong opinion about the

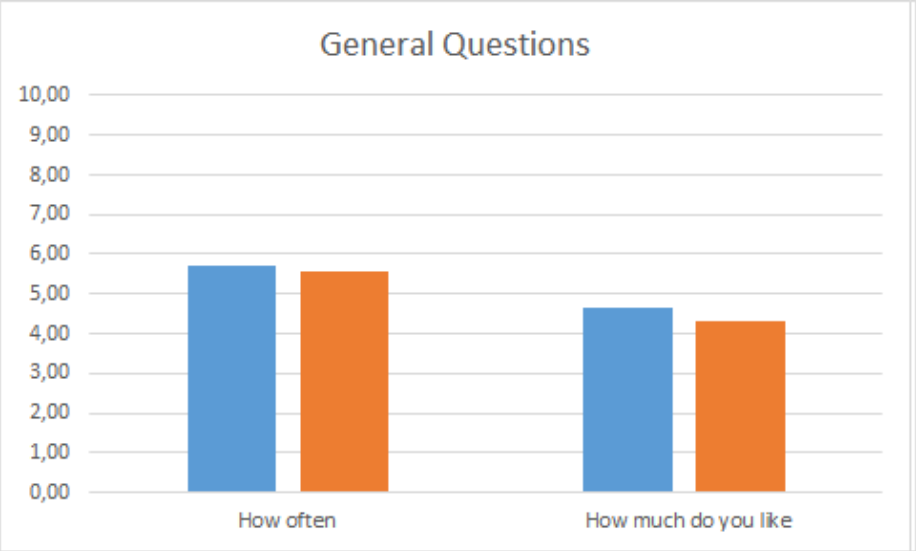

Figure 2: The average of the answers to the general questions.

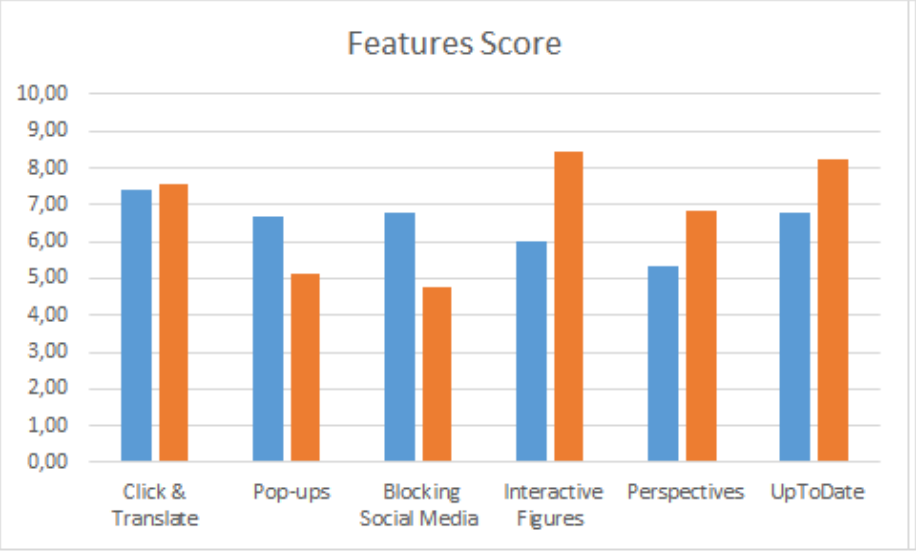

Figure 3: The average of the answers to the feature questions.

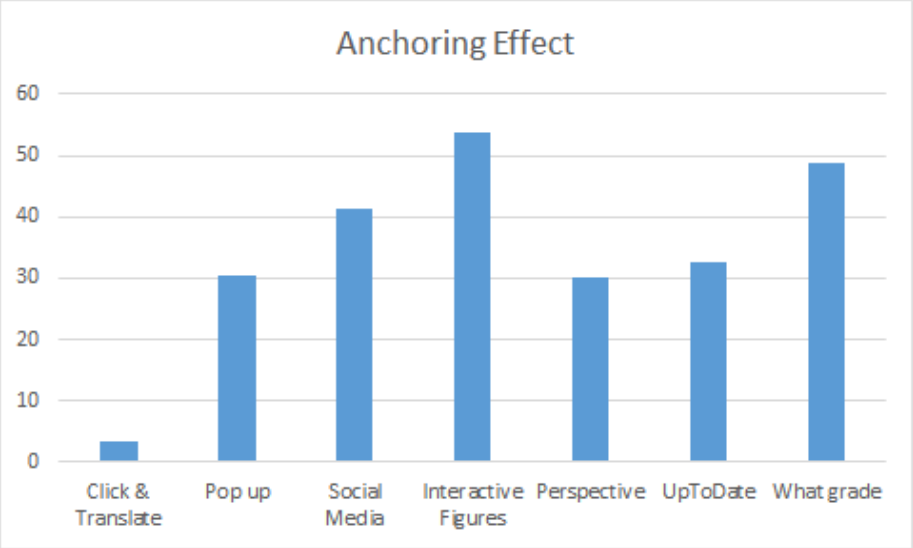

Figure 4: Anchoring index for each feature.

prototype they might adjust towards the anchor and the results will not be entirely true.

4.2.2 Group Dynamics & Strong Opinions

Figure 4 shows that the anchoring effect is not the same in all the questions. We noticed by talking with some participants that they are more drawn to the anchor when they do not have a strong opinion about a feature. This could be explained by group dynamics. The participant prefers to follow the average, the opinion of the group. On the other hand, when participants are strongly opinionated on a subject, we noticed that they also get affected by the anchor but in the opposite way. Instead of changing their answer towards the anchor they drive it away. When we asked why they did that their answer was that they thought that the rate of this specific feature was really unfair so they wanted to change the average towards their opinion.

4.2.3 Availability Heuristic

The last question of the survey was asking the participants to rate the whole eTextbook. In this question we noticed that the survey that ended with the bad results and was anchored also lower had lower ratings, whereas the other survey had a higher rating (see Figure 5). We believe that this result is a combination of the anchoring and the availability heuristic since the participants could easier recall the last features.

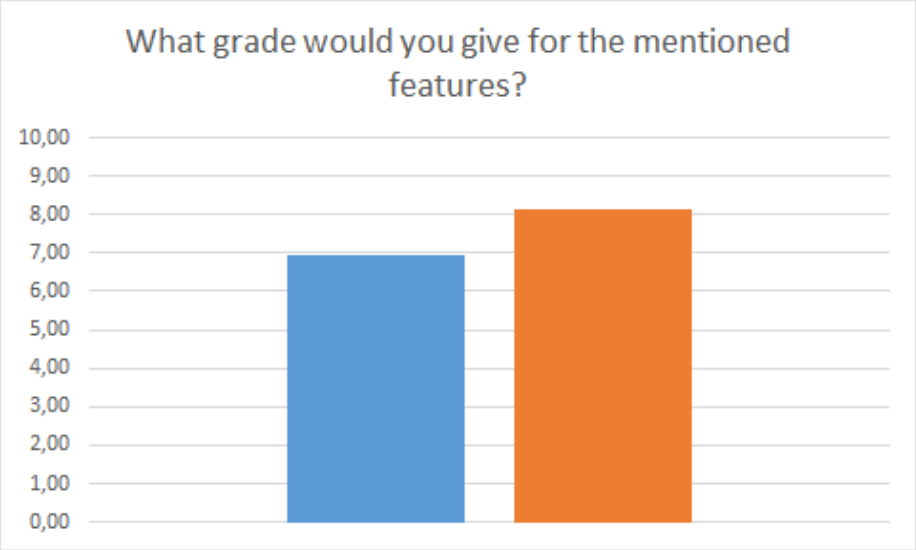

Figure 5: Overall rating of the features.

4.3 Conclusion

The conclusions of this experiment verified the following guidelines. In order to eliminate the influence from group dynamics the participants should not be able to exchange opinions during the prototype session. Additionally, information that indicates current rating or opinions about the prototype should not be given to participants to avoid anchoring.

5 Group Dynamics Experiment

This section contains the group dynamics expirement. In the first subsection the setup of the expirement is explained in detail. The results of the expirement are given the second subsection. The third subsection contains the analysis of these results and fourth and final section provides the threads that are embedded in this experiment.

5.1 Setup

This experiment focusses on group dynamics; the effect of a group on fault detection. The anchoring effect experiment as described in 4 used a 'current score' field to display a fake average of what people voted for the feature. Since this value implies that other people don't like it, participants did not want to deviate to far from the rest and anchored their score to the shown value.
To further investigate the effect of group dynamics, this experiment consists of 2

parts. The first part focuses on groups of people. In the group we ask for them to find good and bad points of the shown feature. The second part asks the same but on individuals. The amount of found points (good and bad) of the group is then compared to the points found by the individuals.

Theory and Expected Outcome The paper from [BarnLund [1]] states "group members saw different issues and a larger number of issues than a single person did working alone". This is because group members will inspire each other with ideas they did not thought of before.
Therefore we expect the following result:

- The group finds a greater amount of points than the individuals because each participants' points inspires the other.

However, we expect the group members will use the inspiration to build up on the consensus of the group. According to [Janis [2]] group members put the value of being part of the group higher than anything else. They will strive for unanimity on issues the group has to deal with. The group members will follow the opinion of others especially the one that takes the lead. For our experiment this will mean that they will follow the group member that is the first one to come up with an opinion. The other group members will build on that opinion and not consider other options. This is also stated in [Jenness [3]], where it is called the "impression of the universality". Agreement becomes the criterion of correctness, meaning that if everyone agrees, which they will because of what was previously said, the statement is a justified true belief by consensus.
This leads us to the following hypotheses:

- The group finds more good or bad points depending on which type of point they start with (if one finds the feature bad, the other participants don't want to stand out and find more ways it is indeed a bad feature). The individuals find a more balanced amount of good and bad points.

5.2 Results

Feature: Pop up

- Group A (4 Persons)
  - Really annoying
  - If you read concentrated it is distracting
  - If placed more subtle (like in the bottom, or as a simple link), it could be really useful
- Group B (3 Persons)
  - Really distracting, advertisement
  - Already implemented as notes on the bottom of the page

  - Should be able to disable it

- Individuals
  - Why would it require to pop-up? Can't it be in the sidebar
  - Can see that it could be useful to add additional content
  - Can point out important info so you can remember it better.
  - Possibly pop-up spam when you scroll through.
  - Extra information helps learning.
  - Can be a distraction.
  - Only useful if a teacher does it, author should just put some extra facts or a note in the text
  - Helps with learning
  - Annoying if there are multiple pop-ups at the same time.
  - Helps with putting your attention to what is important.
  - Can be useful the first time you scan the text but then you should be able to disable it

Feature: Graph

- Group A (4 Persons)
  - Interesting to adjust the parameters with sliders when you can observe the values
  - Too heavy or difficult for current e-readers

- Group B (3 Persons)
  - Nice way to interact if you can see the changes on the values
  - Would it change learning experience (concern)
  - Already existent in executable books.

- Individuals
  - Looks cool
  - How do you use this? Am I lacking math skills? Does it update the math function in the e-book?
  - Real-time feedback
  - Easier to see relations between values and representation
  - Interaction helps understanding.
  - Helps to understand the behaviour of a formula
  - Should not be able to do this if it refers research results / statistics, they should not be changed.
  - Can be useful with difficult formulas, see what happens when values are added.

Feature: Perspective

- Group A (4 Persons)
  - Definition view really good for the exam
  - Picture view less important
  - You can not put every story in a picture
  - Extra work for writers
- Group B (3 Persons)
  - Extra work for the writers
  - Lose connections by seeing only one view
  - Nice addition, could be useful depending on what you want to do with the book
- Individuals
  - Can imagine you want text (the e-book text or some story) visible using different views
  - Definition view could provide more detail and image view more a global idea.
  - Freedom to choose your reading style.
  - Different views creates a diverse and less boring experience.
  - Natural in use; you use different representations for different things.
  - Maybe not all information can be properly defined in both formats.
  - Swapping between views could make reading hard/cluttered.
  - Helps with different learning styles
  - You have to be able to swap between perspectives at any time.
  - Mix views together
  - Easy to adjust to my reading style.

5.3 Observations and Analysis

In this section the results from the expirement are discussed. First the observations on both the group as well as the individuals are stated. The final subsection contains the analysis of the differences of both these groups.

5.3.1 Group Observations

In the group prototype sessions we observed how group dynamics affected the direction of the discussion.Our first observation was that when a feature was presented and one of the participants had a strong negative opinion about it then the other participants that were unsure followed this opinion and it was impossible to change the group effect even if we started indicating some good points of this feature.

The next observation considers the speed with which the group arrives at a consensus depending on the starting idea. If this idea is perceived by the group as good, then everyone agreed and there was no further discussion, we had to doubt this idea to make the group start again the discussion.

The final observation regards how the group builds their perception of an idea on the first idea that comes up in the discussion. Even if the idea does not have a solid base, the other participants will add their ideas on the first one without refuting or doubting it. This is called group thinking [2].

5.3.2 Individual Observations

The first thing that was observed was that the individuals don't come up with new information after the first few points. Since they don't receive information or input from other people they are unable to create new pros and cons as they can only work with what they have in their own mind.
Another observation was noticeable once asking several individuals; since there is no communication between the individuals, they did not mention the same pros and cons. Since they don't communicate there is no influence on each other so their thought process stays there own; their thought processes won't 'synchronize' and results in different ideas per person.
The individuals also asked questions if they were unsure. For example with the perspective: can it also swap?

5.3.3 Analysis of the results

5.4 Threats to validity

Our experiment focusses on group dynamics and individuals. This section lists possible threats of invalidating our results. The results, of both session types, could be affected by our presence and attention.

5.4.1 Time of observation

For group discussions, we handed over our laptop with the displayed feature. So the first person usually talked about what was seen, while the others had not yet viewed our image yet. This could perhaps generate further bias for those who had not seen the feature yet, as they had no way to generate an honest

opinion. We tried to make sure our subjects did not speak loudly about the feature before everyone saw it.

5.4.2 Group dynamics

In the group discussion a number of points were directed towards us instead of the fellow participants. These lead to less open discussions and more about asking validation questions: "Does this mean..." or "How does this work..". We tried to keep ourselves out of discussion which worked sometimes however the less open discussions are a thread as group discussion and group dynamics changed based on our responses.

5.4.3 Individual time

The time spent by individuals was more than those of the groups combined. The individuals felt the obligation to participate and therefore spent more time on the expirement than the groups where no-one felt obligated to talk (Bystander effect).

5.4.4 Proceeding the discussion

Similar to group dynamics, but also in the individuals, the discussion/thought process sometimes felt silent. In order to keep to the discussion going we sometimes restated the question. This lead to new ideas in both the groups and the individuals.

References

[1] A comparative study of individual, majority, and group judgment. Barnlund, Dean C. The Journal of Abnormal and Social Psychology, Vol 58(1), Jan 1959, 55-60. doi:

[2] Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. Janis, Irving L. Oxford, England: Houghton Mifflin. (1972). viii 277 pp.

[3] Jenness, A. Journal of Abnormal and Social Psychology, 1932, Vol.27(3), pp.279-296 [Peer Reviewed Journal] SciVerse ScienceDirect Journals