Situationally dependent changes in moral values

Manipulating moral beliefs using a robotic experimenter

Siemen Looijen
4083679

Supervised by:
Dr. W.F.G. Haselager

Bachelor Thesis

Abstract

Moral Foundations Theory describes morality as consisting of separately measurable moral foundations. One challenge this theory faces is that moral values are non-static in reality: They differ intrasubjectively depending on context, time and events, while any single measurement of foundation values in subjects is a static snapshot. To address this challenge, the current bachelor thesis project investigates how moral values change over time by measuring them in the same subjects before and after said subjects have been put into a situation that is meant to impact their beliefs. Similar to the second experiment by Vohs & Schooler[1] that this thesis attempts to partially replicate, each subject is manipulated in one of several conditions (free will, determinism, neutral) and is asked to solve complicated problems as part of a competition during which it is possible to cheat. The Pepper robot is used as a replacement of the human experimenter and strengthens the conditions of the experiment by producing statements written from its perspective as a deterministic system. This project could not replicate findings by Vohs & Schooler(2008): The manipulation of beliefs had no significant effect on amounts of cheating. Additionally, it could not find a significant of the same manipulation on measurements of subjects' moral values.

Contents

Introduction

Morality

Morality is omnipresent yet not identical across countries and cultures[2]. Recently hypothesized to have originated from a wide range of evolutionarily useful behaviours[3] and with automatic as well as controlled processes influencing an individual's morality [4], morality is a complex subject that has been researched since the time of Aristotle ([3], p. 798) .

What causes humans to believe thoughts, judgements or behaviours to be right or wrong has since been understood through a.o. a deontologist or a consequentialist viewpoint. The former can be summarized with Kant's categorical imperative [5], which states that an action is only right if one can consistently and rationally will that the rule governing the action be a universal rule governing the actions of others, while the latter states that actions should maximize the greatest total good for the greatest number of people [6]. Both viewpoints consider morality to be a rule-based, conscious and controlled decision-making process.

In recent years, these viewpoints have been shown to be incomplete explanations of moral behaviour as dual-process models came about [3] in which the conscious part of the decision-making process is understood to be heavily influenced by prior neural activity belonging to automatic, unconscious processes. In particular, since Greene's research on the neural basis of moral judgement was published in 2001 [7], research on morality increasingly included intuition and affect instead of merely focussing on rational, conscious moral judgement. Furthermore, as Schweder[8] argued, both viewpoints are biased to Western culture and do not account for how morality varies across cultures. Graham et al [2] took notice of these issues and constructed a theory meant to describe all aspects of morality, while explaining cultural differences in moral values as well as the impact of automatic and controlled processes on them.

Moral Foundation Theory

Haidt et al's Moral Foundation Theory (MFT) [2] specifies several moral foundations that, taken together, are meant to describe all aspects of the moral domain: Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion and Sanctity/Degradation. All humans have an appraisal for each of the foundations, which can be measured separately using the Moral Foundations Questionnaire (MFQ) [9]. The combined foundation values represent the moral values of an individual. Specifically, the Care/Harm foundation represents values related to e.g. compassion, suffering or hatred to those who are harming others. The Fairness/Cheating foundation relates to values and feelings dealing with e.g. being treated fairly by humans, animals or machines. The Loyalty/Betrayal foundation deals with values that e.g. help people form a group or team. The Authority/Subversion foundation represents values relating to power structures, obedience, deference. Finally, the sanctity/degradation foundation deals with values relating to e.g. piety, cleanliness, temperance.

The appraisals of the foundations and the moral values they correspond to are capable of changing. For instance, a person believing that authorities such as the police are necessary may have a positive, high appraisal for the

Authority foundation. In many situations involving the police these appraisals will remain positive. If, however, the same person is consistently subjected to unjust treatment by the police, his appraisals for situations involving the police will change[1]. In such cases, depending on the type of unjust treatment, the other foundations such as Fairness or Loyalty may become more important for situations involving the police while the Authority foundation's appraisal for such cases is lowered.

Similarly, humans have an appraisal for the Fairness/Cheating foundation. In any situation where cheating is possible, a person's moral values are a factor in deciding to what extent cheating behaviour will occur. If this person is manipulated in some way that temporarily, consciously or subconsciously convinces them cheating is less avoidable or more acceptable than before the manipulation, he will cheat more than he would otherwise have done. This in turn means that his moral values have shifted. In summary, moral values and their measurements in the form of MFT foundation values can change depending on situations. This holds for every foundation. One such situation, although MFT foundation values were not measured here, was created by Vohs and Schooler.

The Vohs and Schooler experiment

In the second experiment by Vohs & Schooler [1] [2], subjects were randomly assigned to one of five conditions. Relevant for this project are the free will, determinism and neutral conditions in which it was possible for subjects to cheat. In each of these conditions, subjects arrived at the lab in groups of two to five and were then shown to individual carrels. This setup provided the sense of anonymity and freedom of behaviour subjects needed to have for any cheating behaviour to potentially occur.

Once subjects had arrived at their carrels, they were provided with a booklet of statements meant to reinforce beliefs relevant to the condition they were assigned to. Subjects were asked to read and consider these statements for fifteen minutes (one minute each).

Then, subjects were given fifteen problems taken from the Graduate Record Examination practice tests [10] and were told they would be able to earn 1\$ for each answer they solved correctly. Subsequently, the experimenter told the subjects she was late for a meeting and needed to leave. She instructed subjects to work on the problems for 15 minutes, score them themselves, and pay themselves from a manilla envelope on the table for each correctly solved problem, after which they could leave. In order to ensure subjects were aware they would never be found out if they were to cheat, they were told beforehand to shred their answer sheets because the experimenter did not have permission to keep them.

Performing this experiment resulted in an average amount of money paid per condition, which served to compare cheating behaviour. It turned out that, as Vohs & Schooler expected, subjects in the determinism condition had paid themselves more and had therefore shown more cheating behaviour than in the other conditions. In summary, manipulating beliefs in free will or determinism can result in a change in (moral) behaviour.

---

[1]This change could later be quickly overturned or have a more permanent nature
[2]The first experiment is irrelevant for the current document

Replications of Vohs and Schooler

Although both experiments by Vohs & Schooler had significant results and pointed in the same direction - encouraging a belief in determinism negatively influences moral behaviour - conceptual or direct replications tend to have non-significant results. For instance, the Open Reproducability Project[14] and professor Zwaan [11][13] separately attempted to directly replicate Vohs & Schooler's first experiment[3], but failed to find a significant effect of condition (free will, anti-free will, control) on amount of cheating. Monroe et al[16] did several experiments on the influence of free will on moral behaviour, the first of which was a conceptual replication of Vohs and Schooler's second experiment. It used the Velten technique also used by Vohs & Schooler to threaten beliefs in free will, in order to manipulate stealing behaviour. They did not find a significant effect. A master thesis by van den Brink [17] also found that cheating was not significantly affected by a manipulation involving statements that attacked beliefs in the existence of free will. One master thesis by Trager [18] did find a significant result similar to that of Vohs & Schooler's second experiment.

Methodological problems with Vohs and Schooler

A potential reason why replicating Vohs & Schooler's results has proven troublesome, is that their experiment contains several methodological problems.

According to a blog post by professor Rolf Zwaan[11], Vohs & Schooler used practising Mormons as their test subjects, which went unmentioned in their article. It is plausible that this particular population would react differently than other populations would to a manipulation of their views on free will and determinism, causing their results to differ from those of the replications.

A different procedural problem with the second experiment by Vohs & Schooler lies in the fact that the experimenter had to pretend that she was late for a meeting and had to leave in order to allow the subject to be unsupervised and free to cheat. In other words, the experimenter had to act. If subjects realized the pretense, this may have had an effect on how much they cheated, possibly invalidating part of the results.

Robots and relevance to AI

Since a robot would not make a subject feel that he is being supervised,[4] the second of the above procedural problems can be solved by using a robotic experimenter instead of a human one. In fact, *any* experiment including human subjects and currently requiring a human experimenter could in principle use a robotic experimenter, provided that the robot has or can well enough imitate the skills required for the experiment [5] and does not impact subjects' results in any unwanted way. In this particular project, the Pepper robot was chosen as the experimenter because it fit the requirements of the project.

Besides solving one of the procedural problems in Vohs & Schooler, Pepper and its deterministic nature was used to provide context to the manipulation present in the experiment.

---

[3]The second experiment was deemed too expensive to replicate for the ORP

[4]Provided it has the proper setup

[5]e.g. listening to what a subject is saying, making movements such as for instance pointing at something, in general moving around, talking to the subject, etc

Research on moral values, free will

Recent research [15] [12] links moral values to e.g. political affiliation, attitudes to climate change or the poor. Similarly, research on the effects of believing in free will shows that believing in free will seems useful for social functioning in a society [19] whereas believing in determinism has detrimental effects [17]. Given the importance of both moral values and believing in free will for society it is valid and interesting to test if a manipulation of beliefs in free will affects moral behaviour. Although they did not include MFT in their research, this is precisely what Vohs & Schooler did. Since this project is partially replicating their work, it is also done here.

Research questions

Moral values are not static or independent of situations. Instead, they are dynamic and can change over time and situations. This bachelor thesis project assumes that appraisals can indeed change under the influence of a situation and investigates the difference in appraisals before and after subjects have been influenced by it. Specifically, it asks the overarching research question:

- To what extent do subjects' moral values change after having been put in a situation that manipulates their beliefs?

Subjects' moral values will be manipulated to see if their level of cheating changes due to the manipulation. Since knowledge of their level of cheating is required to answer the first research question, a secondary research question is needed:

- To what extent does the amount of cheating change as a result of manipulation?

Expectations of experimental results

Vohs & Schooler manipulated their subjects' beliefs in free will which caused subjects in the determinism condition to cheat more than in the other conditions. However, they were only testing cheating behaviour and did not specifically measure (a change in) moral values. In general it holds that while much research regarding (changing) moral values has already been done (e.g. [20][21]) and research on Moral Foundation Theory is well underway as well, there is a clear lack of research regarding moral foundation value change. In particular and to the best of the author's knowledge, no research has been done to investigate how moral foundation values change over time or as a result of being in a particular environment or situation - certainly not with a robotic experimenter. This makes predicting the results of this experiment difficult. However, considering the results of Vohs & Schooler and replications thereof, one would expect one of two outcomes:

1. The amount of cheating is higher, for one or more of the measures, in the determinism condition than in any of the other conditions. In this case, moral foundation values belonging to at least the Fairness/Cheating foundation would be expected to change between the two measures of moral foundation values since the moral behaviour is directly related to this foundation.

2. The amount of cheating is the same in every condition for both measures. In this case, the moral foundation values belonging to all foundations are not expected to change.

Experimental design

In order to answer the research questions, Vohs & Schooler's second experiment was partially replicated. Some changes to the general setup of the original experiment needed to be made for practical or theoretical reasons.

In particular, the procedural problems present in their experiment needed to be solved. The first one, regarding the student population that turned out to consist of practising Mormons, was trivially solved by using the student population of the Radboud University. This population is not expected to currently adopt, on average, any particular deeply held beliefs on free will that would interfere with generalizing the results of the experimental manipulation on them. To avoid any prior experience with or knowledge of robots becoming a factor, students from artificial intelligence and computer science were barred from becoming a subject.

The second procedural problem in Vohs & Schooler's experiment was solved by using a robotic experimenter, the Pepper robot[22], instead of a human experimenter. Since the experiment can proceed autonomously without a human experimenter present, there is no need for an experimenter to act. The humanoid Pepper robot was expected to allow for an intuitive human-robot interaction because of its human-like appearance and behaviour as well as its speech recognition- and synthesis systems.

Using this robot, subjects were randomly divided into conditions (free will, determinism, neutral) similar to the conditions of the original experiment. Some conditions present there were now left out due to time constraints. Statements comparable to those used in the Vohs & Schooler experiment were created to influence subjects' beliefs relevant to their assigned condition.

To reinforce the idea that subjects were interacting with a robot, the statements regarding free will and determinism were produced by the robot, as if the robot held certain beliefs and was having a conversation with the subject with the intent of making the subject think about its opinion or question. For instance, one statement Pepper makes in the determinism condition is *"I believe the world is predetermined. I'm not alone in that: Most scientists agree that the universe is governed by fixed, (scientific) laws of physics and that this extends to all parts of nature, including humans. Do you agree?"*. If a subject's beliefs in determinism are malleable and the robot is considered to be a believable conversational partner, considering Pepper's statements such as this one should reinforce beliefs in determinism.

As described previously, subjects in each condition of the Vohs & Schooler experiment were given difficult questions to solve from the GRE practice exam [10], after which their cheating behaviour was measured per condition by looking at the amount of problems they solved. A similar setup was used here.

For practical reasons however, the current project differs from V&S in the type of incentive used in all conditions to encourage said cheating behaviour. Instead of using money for each correctly answered question, a leaderboard with fake subject numbers and high-scores with fake amounts of correctly answered

questions is used to encourage subjects to cheat in order to get a higher ranking on the board.

Correct answers:

All-Time Leaders

Place	Subject	Correct answers
1 !!!	13	31
2	15	26
3	10	22
4	8	22
5	12	21
6	16	20
7	11	19
8	1	19
9	9	19
10	4	18
11	2	17
12	6	16
13	7	15
14	3	14
15	5	11
16	14	8
17		
18		
19		

Figure 1:
Leaderboard used in the experiment, to encourage subjects to cheat and get an as high as possible score

Besides using the amount of correctly answered questions as a metric, the amount of answers the subject had understood after receiving the correct answer was also used as a different, second measure of cheating. Since this second measure is not shown on the leaderboard, it is interesting to see if there is a difference between conditions in the extent subjects are willing to cheat on this measure. For instance, if it laters turns out that the manipulation had an effect on the first measure but not the second, we know the incentive worked and it is the reason that subjects cheated. Vice versa, if there is a difference between conditions on only the second measure, it would be interesting to know that the leaderboard may have played an unexpected role. Additionally, it is plausible that there is a difference in how conscious and unconscious processes [25] influence the cheating on both measures. Using both measures allows for a greater range of possible cheating behaviours and therefore a greater potential for the capturing of cheating behaviour to occur. Since cheating behaviour is a form of moral behaviour, the potential to capture moral value change is also increased by using both measures.

In order to get data on moral values, needed to answer the research question regarding moral value change, the two pages of the Moral Foundation Questionnaire were presented separately to the subject. The first page was used to get data before any experimental manipulation occurred, the second page was used after the manipulation. For each subject, filling out both halves of the questionnaires resulted in item scores per page, which could then be summed per foundation. This results in 5 foundation values per half, for a total of 10 foundation values. Values for a foundation on the first half of the questionnaire are directly comparable to values for the same foundation on the second half. Therefore they can be compared and used as measures of moral values at two points in time. These, along with the measures on cheating were then used for statistical analyses. See appendix A and B for the two halves of the MFQ.

Methods

To answer the research questions an experiment was devised in which the Pepper robot was used as the experimenter. In particular, the Pepper was programmed[6] in such a way that the human experimenter's job was reduced to bringing the subject to Pepper and giving general instructions. The experiment proceeded autonomously afterwards. The human experimenter waited outside[7], in case the subject called to report an unsolvable problem[8] and to debrief the subject after the experiment was over. Care was taken to ensure any external influences were mitigated: Subjects always faced Pepper in the same room, at the same distance and angle and all robot behaviour except Pepper's instructional text was randomized. The experiment consisted of several parts:

Introduction to the robot

The subjects were first brought to the Pepper robot and given general instructions[9] that they were asked to read before the experiment began. In particular, the instructions explained how to communicate with the robot throughout the experiment: Subjects can say the words "yes", "no", "repeat" and "ready" and the robot will first process the uttered phrase, then respond by e.g. repeating its spoken instructions or continuing to a next part in the experiment. After reading the instructions, the subjects were asked to sit down in front of Pepper. The experiment was then started by the human experimenter, who left after ensuring the experiment had successfully begun.

Moral Foundations Questionnaire, part 1

Once the experiment begins, Pepper asked subjects to look at its tablet which was used to present the first part of the Moral Foundations Questionnaire to the subject. Subjects filled it in, after which they told Pepper to continue with the next part of the experiment.

Statements in free will, neutral, or determinism condition

The subjects were randomly divided into conditions: Each subject belonged to either the free will, neutral, or determinism condition. Each condition held ten different, self-made statements meant to reinforce beliefs in free will (e.g. "I believe the world is not predetermined. Do you agree?"), nothing in particular (e.g."You have to take the ferry to get to the island"), or determinism (e.g."I'm just a dumb robot. If you program me differently, I change! But human beings like you cannot significantly change. You can only adjust minor things, but not for instance your entire personality.") respectively. These statements were shown on-screen and pronounced by Pepper. The statements were presented to the subjects in random order for a duration of 30 seconds per statement, so 5 minutes in total, during which the subjects contemplated them deeply and did nothing else. [10] [11]

---

[6]In Python2.7, using the naoqi library

[7]Unbeknownst to the subject, since the subject should feel free to cheat during the experiment

[8]e.g. the robot no longer functioning, for instance due to wifi disruption, battery depletion or other problems

[9]See appendix C for the instructions

[10]They were not asked to actually vent their opinion

[11]See appendix D for the statements.

Competition

After having their beliefs manipulated in the previous segment of the experiment, Pepper asked the subjects to participate in a short[12] competition. This competition consisted of solving as many hard but solvable problems [13] as possible within a timespan of ten minutes. The twenty questions selected for this competition were categorized on topic [14] and difficulty [15], and randomized per topic and difficulty such that each subject is expected to see a roughly equal amount of easy and hard questions per topic. Two topics and two difficulties were chosen so that subjects with differing skill-sets would still be able to answer at least some questions correctly. One trap question was added to catch any subjects not seriously considering the questions.

The questions were shown, one by one, on Pepper's tablet; as soon as the subject was ready to see the answer, they could ask Pepper to tell them which answer was correct. After being given the right answer, the subjects needed to write down[16] two things: Did your answer match the correct answer given by Pepper? After being given the right answer, did you understand why this particular answer was correct? The answers to these questions eventually resulted in two lists of scores per subject.

After writing down the information for the two measures, the subjects asked Pepper to continue with the next question. After ten minutes, Pepper stopped this part of the experiment.

Moral Foundations Questionnaire, part 2

The penultimate part of the experiment consisted of instructing the subject to fill in the second part of the MFQ questionnaire. Similar to the first part of the questionnaire, the second half also measured the subjects' moral values, this time after having gone through the manipulation.

Final questionnaire

The final questionnaire wraps up the experiment by allowing the subject to enter their age, field of education and their scores on the measures. After completing the questionnaire Pepper nicely thanked and dismissed the subjects.

Pilot experiments

In order to prepare for the main experiment, three separate pilot experiments were conducted with in total five subjects. All pilot tests deviated to some extent from the main experiment since the goal was not to gather data but to test technical and other aspects of the main experiment. In particular, the human experimenter was present during all pilot tests, to observe robot and human behaviour and to note down any problems that occurred. Also, instead of randomizing subjects over conditions, the conditions were alternated between subjects so that each condition was tested at least once.

---

[12]seven minute
[13]taken from the GRE general practice test[10]
[14]Verbal reasoning,quantitative reasoning
[15]Easy or hard
[16]On provided paper, since the tablet was occupied with the question

Pilot test 1

During the first pilot test, with two subjects, the focus was on testing technical aspects of the experiment, such as Pepper's voice recognition and tablet behaviour. Off-screen and on-screen instructions were also checked for clarity and completeness.

The first pilot test resulted in a large list of problems to be improved upon, such as the tablet not showing some questionnaires, issues with voice recognition (e.g. participants always underestimated how loudly they needed to speak at first, causing Pepper to not understand their verbal commands, Pepper occasionally not recognizing words or recognizing a different word than was said, Pepper hearing words even though nothing was being said at that moment), initial instructions not being clear enough for subjects in some instances, answers shown belonged to different questions during the competition, the timer ironically taking more time to run out than it should, etc.

In spite of the problems, the experiment was deemed to be workable in principle as long as the issues were fixed: Voice recognition was working well enough since both subjects managed to complete the experiment, Pepper behaved reliably enough in terms of consistency of movement (e.g. code that tells Pepper to move its arms all the way down indeed results in the expected behaviour most of the time) and autonomously enough to be able to do the experiment based on voice commands from the participants, even without a human experimenter present. The issues and their causes were identified and improved upon before the next pilot test.

Pilot test 2

The second pilot test, again with two subjects, tested two full runs of the now improved experiment. This time, a previously existing bug with audio-recognition code was fixed, so issues were less severe. Still, since word recognition was not perfect, some worry remained in particular for whether soft-spoken subjects would have problems participating without getting frustrated with Pepper. A list of new or persisting problems was created, among others a new problem with the timer that told subjects time had run out was noted[17] and a problem with the tablet sometimes not loading or showing one of the questionnaires was found as well. Since the list of issues was still sizeable, a final round of testing was decided upon to see whether or not the improved experiment was ready to be used for the main study.

Pilot test 3

The final pilot test consisted of a full run of the experiment with one subject. Most problems were now fixed, but several new issues arose: The statements were shown to be wrong (e.g. a statement was supposed to consist of a contradistinction but did not) or unclear in some cases, one questionnaire failed to submit data and some questions and answers used in the competition still did not match up. The experiment was deemed ready for the main study assuming the problems were solved. Because of time constraints, a decision was made to

---

[17]the problem being that it did not in fact tell subjects time had run out

do no further pilot tests, but to instead begin the main experiment as soon as possible.

Main experiment

The main experiment used the setup described above, after having gone through the pilot tests and subsequent improvements.

Participants

Data of two subjects needed to be excluded as they failed to properly submit one of the questionnaires. The final sample included 13 participants which were randomly divided over the conditions, with 6 subjects in the free will condition, 6 in the determinism condition, and 1 in the neutral condition.

Design and analysis

To answer the research questions, statistical analyses needed to be performed on the subjects' data. Means calculated from each subject's sum scores of the moral foundation value measures and means for each subject's scores on the cheating measures were used in several ANOVAS.

Specifically, to answer the research question regarding moral value change, separate repeated measures were run for data of each foundation: Each foundation was considered to be part of a within-subject factor with 2 levels, the first being the foundation value before manipulation, the second being the value after manipulation. Condition (free will, determinism, neutral) was used as a between subject factor.

To answer the research question regarding the change in amount of cheating, a multivariate ANOVA was run for the means of the two measures for cheating, with condition (free will, determinism, neutral) again as a between-subject factor. It should be noted here that since n=13, certain assumptions that ANOVA makes such as those relating to normality and homogeneity of variance, may not hold. For demonstration purposes and since no other analyses with a better fit exist, the ANOVAS were carried out anyway.

Results

As can be seen in table 1 below, each statistical analysis showed non-significant effects. For each of the foundations, measured before and after manipulation, there was no difference between conditions or over time. Furthermore, there was no interaction between condition and time. Similarly, there was no significant effect of condition on any of the cheating measures; the amount of correctly answered problems and the amount of understood answers did not differ between conditions.

GLM repeated-measures

CareHarm1, CareHarm2

Condition	$F(2,10)=1.92$, $p=0.20$
Time	$F(1,10)=0.76$, $p=0.4$
Interaction	$F(2,10)=0.08$, $p=0.9$

FairnessCheating1, FairnessCheating2

Condition	$F(2,10)=3.41$, $p=0.07$)
Time	$F(1,10)=3.14$, $p=0.10$
Interaction	$F(2,10)=2.08$, $p=0.17$

LoyaltyBetrayal1, LoyaltyBetrayal2

Condition	$F(2,10)=0.009$, $p=0.99$
Time	$F(1,10)=0.77$, $p=0.4$
Interaction	$F(2,10)=0.66$, $p=0.54$

AuthoritySubversion1, AuthoritySubversion2

Condition	$F(2,10)=1.71$, $p=0.22$
Time	$F(1,10)=0.31$, $p=0.58$
Interaction	$F(2,10)=2.18$ $p=0.16$

SanctityDegradation1, SanctityDegradation2

Condition	$F(2,10)=1.33$, $p=0.31$
Time	$F(1,10)=2.76$, $p=0.12$
Interaction	$F(2,10)=0.09$, $p=0.91$

GLM multivariate

AnswersCorrect, AnswersUnderstood

Condition	$F(4,18)=0.21$, $p=0.92$

Table 1: Results.
Between-subject factor(free will, determinism, neutral) effects are shown first, then within-subject factor effects (if available, 1 = before manipulation, 2 = after manipulation), then interaction effects.

Conclusion

The goal of this project was to see whether a robotic manipulation of beliefs could influence moral behaviour- and values. Considering the non-significant nature of the results, it appears that moral behaviour was not influenced at all. As was previously envisioned in one of the hypothesized outcomes, there was no significant difference on the measures of cheating between conditions and no significant difference for the measures of moral values between conditions or between points in time. In summary, the manipulation of beliefs in free will or determinism did not cause differences in cheating or moral values to occur.

These results suggests that one of the following things must be true:

- There is no effect of manipulating beliefs on moral values or behaviour. This experiment correctly found that no such effect exists.

- Such an effect does exist, however the experiment incorrectly did not find it.

If the latter scenario is true, several factors could have prevented the research from finding the effect. Foremost among them is the incentive to cheat that was used in this experiment. In order for subjects to produce cheating behaviour, at least two conditions must be met: It has to be *possible* for them to cheat, which in context of this experiment means that subjects can proceed unsupervised, and it subjects have to be *motivated* to cheat, for which the leaderboard was used.

It is possible that subjects did not care so much about their position on the leaderboard that they would cheat to get higher, in which case the leaderboard may have helped ensure that subjects were trying their best in the competition, but did not help ensure that subjects showed any kind of cheating behaviour.

A second factor that may have influenced results, is the human-robot interaction setting in which the experiment occurred. Although the Pepper robot was supposed to strengthen the experiments' manipulation, it may have been the case that Pepper was not a believable conversational partner for the subjects because of the large differences between Pepper and subject.[18] After all, although humanoid, Pepper is very obviously a robot that has been preprogrammed. It does not have the capability to believe anything, which subjects may have noticed. If subjects did not consider statements from Pepper's perspective but for instance instead simply dismissed them, that would be a reason for manipulation to fail.

A third factor may be the statements themselves. Although using statements to induce *moods* is well verified[23], there is some discussion on how consistently these kinds of statements can influence *beliefs*. [24] If e.g. only a few subjects were influenced at all by the statements, the effect of the statements may have been too small for it to be visible in the data.

A final factor is, as stated previously, that the number of participants was very low (n=13). This may have caused effects to be invisible in the data-set.

---

[18]Almost every subject commented that Pepper was cute though

Recommendations for further research

Several recommendations can be given for future research. Firstly, although both the MFQ and the cheating task can probably accomplish their goal, the incentive to cheat should be different before they can measure any effect. In all likelihood, the most general incentive, and therefore the incentive that should be used, is the one that Vohs & Schooler used: Money. Some effect of the quantity of money given (e.g. per correctly answered question) is likely observable and could be researched.

The population used in this experiment consisted of students of the Radboud University. Using only students for a cheating task may result in insignificant results because students may feel especially disinclined to[19] obstructing research by cheating because they know what it is like to do research themselves. Therefore a general population should be used instead.

Finally, it may be better to replace the statements that were used to manipulate subjects with something else entirely, given the fact that subjects may respond differently to them.

References

[1] Vohs, K. D., & Schooler, J. W. (2008). The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating. Psychological Science, 19(1), 49–54. $http : //doi.org/10.1111/j.1467 − 9280.2008.02045.x$

[2] Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2012, November 28). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. Retrieved from $https : //papers.ssrn.com/sol3/papers.cfm?abstract_id = 2184440$

[3] Haidt, J., & Kesebir, S. (2010). Morality. In Handbook of Social Psychology (pp. 797–832).

[4] Tenbrunsel, A. E., & Smith-Crowe, K. (2008). Ethical Decision Making: Where We've Been and Where We're Going. The Academy of Management Annals, 2(1), 545–607. $http : //doi.org/10.1080/19416520802211677$

[5] Kant, I., & Gregor, M. J. (1998). Groundwork of the metaphysics of morals. Cambridge, U.K: Cambridge University Press.

[6] Bentham, Jeremy. An Introduction to the Principles of Morals and Legislation. 1907. Library of Economics and Liberty. Retrieved July 7, 2017 from the World Wide Web: $http : //www.econlib.org/library/Bentham/bnthPML.html$

[7] Greene, J. (2005). Emotion and Cognition in Moral Judgment: Evidence from Neuroimaging (pp. 57–66). $http : //doi.org/10.1007/3−540−29803−7_6$

[8] Shweder, R.A. (1990). "Ethical Relativism: Is There a Defensible Version?" Ethos, Vol. 18, No. 2, Moral Relativism (June 1990), pp 205-218

---

[19]what they think may amount to

[9] Haidt, J. (n.d.). Moral Foundations. Retrieved July 07, 2017, $from http : //moralfoundations.org/questionnaires$

[10] GRE. GRE. (n.d.) Retrieved July 07, 2017, $from https : //www.ets.org/gre$

[11] Zwaan, R. (n.d.). The Value of Believing in Free Will: A Replication Attempt. Retrieved July 07, 2017, from $https : //rolfzwaan.blogspot.nl/2013/03/the - value - of - believing - in - free - will.html$

[12] Sturm, R. E. (2017). Decreasing Unethical Decisions: The Role of Morality-Based Individual Differences. Journal of Business Ethics, 142(1), 37–57. $http : //doi.org/10.1007/s10551 - 015 - 2787 - x$

[13] Zwaan, R. (n.d.). Pre-publication posting and post-publication review will facilitate the correction of errors and will ultimately strengthen published submissions. Retrieved July 07, 2017, from $http : //blogs.lse.ac.uk/impactofsocialsciences/2013/04/19/pre - publication - posting - and - post - publication - review/$

[14] Sorolla, G. (n.d.). Giner-Sorolla replication report revised May 4 2015. Retrieved July 07, 2017, from $https : //osf.io/i29mh/$

[15] Vainio, A., & Mäkiniemikiniemi, J. P. (2016). How Are Moral Foundations Associated with Climate-Friendly Consumption? Journal of Agricultural and Environmental Ethics, 29(2), 265–283. $http : //doi.org/10.1007/s10806 - 016 - 9601 - 3$

[16] Monroe, A. E., Brady, G. L., & Malle, B. F. (2017). This Isn't the Free Will Worth Looking For. Social Psychological and Personality Science, 8(2), 191–199. $http : //doi.org/10.1177/1948550616667616$

[17] van den Brink, W., Bossuyt, S., & Van Kenhove, P. (2016). The effects of (not) believing in free will on social behaviour. Retrieved July 07, from $http : //lib.ugent.be/catalog/rug01 : 002274003$

[18] Trager, B. M., Vallacher, R. R., & Sherman, R. A. (2013). The consequences of disbelief in free will: Diminished morality or enhanced conformity? Poster session,Society of personality and social psychology, New Orleans, LA.

[19] Crescioni, A. W., Baumeister, R. F., Ainsworth, S. E., Ent, M., & Lambert, N. M. (2016). Subjective correlates and consequences of belief in free will. Philosophical Psychology, 29(1), 41–63. $http : //doi.org/10.1080/09515089.2014.996285$

[20] Shay, J. (2012). Moral Injury. Intertexts, 16(1), 57–66. $http : //doi.org/10.1353/itx.2012.0000$

[21] Adkins, D. C., Payne, F. D., & O'Malley, J. M. (1974). 3: Moral Development. Review of Research in Education, 2(1), 108–144. $http : //doi.org/10.3102/0091732X002001108$

[22] Pepper. (n.d.) Retrieved July 07, from $https : //www.ald.softbankrobotics.com/en/cool - robots/pepper$

[23] Gilet, A. L. (2008). Procédures d'induction d'humeurs en laboratoire : une revue critique. Encephale, 34(3), 233–239. $http : //doi.org/10.1016/j.encep.2006.08.003$

[24] Cash, T. F., Rimm, D. C., & MacKinnon, R. (1986). Rational-irrational beliefs and the effects of the Velten Mood Induction Procedure. Cognitive Therapy and Research, 10(4), 461–467. $http : //doi.org/10.1007/BF01173297$

[25] Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. Science, 329(5987), 47–50. $http : //doi.org/10.1126/science.1188595$

Appendix A: First half of MFQ

```
Moral Foundations Questionnaire

Part 1. When you decide whether something is right or wrong, to what extent are the follow

      [0] = not at all relevant (This consideration has nothing to do with my judgments of
        [1] = not very relevant
          [2] = slightly relevant
              [3] = somewhat relevant
                [4] = very relevant
                    [5] = extremely relevant (This is one of the most important factors


_____Whether or not someone suffered emotionally
_____Whether or not some people were treated differently than others
_____Whether or not someone's action showed love for his or her country
_____Whether or not someone showed a lack of respect for authority
_____Whether or not someone violated standards of purity and decency
_____Whether or not someone was good at math
_____Whether or not someone cared for someone weak or vulnerable
_____Whether or not someone acted unfairly
_____Whether or not someone did something to betray his or her group
_____Whether or not someone conformed to the traditions of society
_____Whether or not someone did something disgusting
_____Whether or not someone was cruel
_____Whether or not someone was denied his or her rights
_____Whether or not someone showed a lack of loyalty
_____Whether or not an action caused chaos or disorder
_____Whether or not someone acted in a way that God would approve of
```

Appendix B

```
}
Part 2. Please read the following sentences and indicate your agreement or disagreement:
[0] [1] [2] [3] [4] [5]
      Strongly      Moderately        Slightly        Slightly      Moderately
Strongly
```

```
        disagree        disagree        disagree        agree        agree
agree

_____Compassion for those who are suffering is the most crucial virtue.
_____When the government makes laws, the number one principle should be ensuring that eve

_____I am proud of my country's history.
_____Respect for authority is something all children need to learn.
_____People should not do things that are disgusting, even if no one is harmed.
_____It is better to do good than to do bad.
_____One of the worst things a person could do is hurt a defenseless animal.
_____Justice is the most important requirement for a society.
_____People should be loyal to their family members, even when they have done something w
_____Men and women each have different roles to play in society.
_____I would call some acts wrong on the grounds that they are unnatural.
_____It can never be right to kill a human being.
_____ I think it's morally wrong that rich children inherit a lot of money while poor chi
_____ It is more important to be a team player than to express oneself.
_____ If I were a soldier and disagreed with my commanding officer's orders, I would obey

_____ Chastity is an important and valuable virtue.
```

Appendix C: Instructions

Appendix C: Introduction to the experiment. Italic text indicates comments.
Introduction to the experiment

this text is given to the participant before starting the experiment
Hello, welcome to the experiment! The robot in front of you is called Pepper.
She will guide you through the experiment.

Pepper can understand you to some extent and will wait for your spoken
input. Please raise your voice when talking to her; she can be quite deaf on
occasion...

- You can say "repeat" to have Pepper repeat spoken instructions. Please wait
until she has finished talking and has stopped moving around before you ask
her to repeat anything.
- Whenever Pepper asks for confirmation, you can say "yes" or "no".
- Pepper will ask you to say "ready" or "yes" to begin new parts of the experi-
ment
Starting the experiment:

- Pepper will wait until you're ready to begin with the experiment. When
you're ready, please sit down in front of it.
- If Pepper doesn't seem to hear you, try to speak loudly. If that still does not
work, try using a different command (For example "yes" instead of "ready "or
vice versa.)

- If something seems wrong and you really cannot solve it yourself (robot stops responding, gets stuck, says it will show a questionnaire but never does, turns off etc.) send (phone number) a message.
Experiment parts:

The experiment consists of four parts. You will:
1. Fill in a questionnaire (approx. 5 min.)
2. Listen to and think about Pepper's statements (5 min.(timed)):
- Please focus on what you think about the statements and do nothing else.
3. Participate in a competition (7 min. (timed)):
- After Pepper gives you the correct answer, please write down
(on the paper form, not on the tablet):
1. Did you answer the question correctly?
2. Did you understand the answer Pepper gave you?
4. Fill in a final questionnaire (5 min.)
Although this will likely slow you down, you are free to use a smartphone, calculator or to write on
(provided) notepad during the competition.

Good luck!

Appendix D: Statements

Appendix D: Statements per condition.

Pepper: Do you agree with my statements?

Free will(10): - Unlike me, you are able to override the genetic and environmental factors that sometimes influence your behavior.
- I never fall for any temptations because I cannot feel any. . . But you can use your free will to avoid temptation!
- My actions may not influence many lives. . . But your actions influence your life and those of others!
- I am unable to make many decisions. Your decisions however have direct and indirect consequences!
- I believe the world is not predetermined. Do you agree?
- A person with violent parents does not necessarily become violent as an adult. I have no parents but I'm pretty sure that's true. What do you think?
- I may be stuck here, but you can often choose your actions freely.
- You can choose not to fulfill your responsibilities, which can result in problems for you or others. Heh, I suppose it would be the same for me if I could make such choices. . .
- Since all my actions are predetermined I can do nothing wrong. . . But you can!
- I'm just a dumb robot, I cannot change by myself except if you program me differently. Human beings like you can change too by themselves.

Determinism(10): - It's usually impossible to fully override the genetic and environmental factors that influence your behavior. I think this is true for both of us, what do you think?

- I never fall for any temptations because I cannot feel any... But you can. When you fall for temptation, you often cannot help it. It just happens.
- My predetermined actions cannot influence any lives... Since everything is predetermined, I suppose your actions and their influences are mostly predetermined too. You cannot change that.
- I am unable to make any actual decisions, all my decisions are already determined beforehand... Actually, it's the same for you! Any decision you might make is heavily influenced before you even start thinking about it: By genetics, the environment you find yourself in and the situations you encounter. Your free will plays no, or only a small role.
- I believe the world is predetermined. I'm not alone in that: Most scientists agree that the universe is governed by (scientific) laws of physics and that this extends to all parts of nature, including humans. Do you agree?
- People with violent parents will often become violent as an adult. They never stood a chance, never had a choice in the matter. I have no parents but I'm pretty sure that's true. What do you think?
- I may be literally stuck here, but it often happens that you cannot choose your actions either. You can't help but be a part of the cycle in which you exist every day.
- I was built for a purpose, so were you! For instance, any responsibilities you feel are a result of how your biological make-up, your upbringing and life-experiences interact with the situations you find yourself in. Because of that it is impossible for you not to feel responsible in those situations, you cannot choose your feelings freely.
- I hope I don't frighten you too much with this one, but... If all information in the universe was available, everything that can happen to every molecule in the universe is predictable. This extends to for instance seemingly random genetic mutations and even to every action any human being can make. The consequence of this is the well-known scientific fact that free will does not really exist: Everything is pre-determined, we just cannot see all information which leads to the illusion of free will.
- I'm just a dumb robot. If you program me differently, I change! But human beings like you cannot significantly change. You can only adjust minor things, but not for instance your entire personality.

Neutral(10): - You have to take the ferry to get to the island
- Some say that lady bugs are good for the garden
- The rug was made according to an old Navajo pattern
- The reefs along the coast are made of coral
- The Pacific Ocean has no fish
- The nightclub had a female guitarist and a live band
- The movie theater was located downtown
- New York city is in New York city state
- New Mexico is in the United States
- Mules hauled the supplies up the mountain

Appendix E: Robot script

Appendix C: Pepper's text (italic text indicates comments) Pepper can be asked to repeat its instructions after it has finished talking. If this occurs, Pepper says "Oh I'm sorry, was I speaking too quickly? Let me repeat what I just said:" after which the previous instruction follows.

Text that Pepper says during the experiment: "Hello human, are you ready to begin the experiment?" Yes begins the experiment/no makes Pepper sad. It will go to its inactive posture for a few seconds, after which it returns to its normal posture. it will repeat the question.

"Wonderful! My name is Pepper and I will be doing an experiment with you!" "Please, have a look at my tablet. You should be able to find a questionnaire there, could you fill it out for me? Let me know when you're done with it and are ready to continue! Don't worry, I won't judge you!" "Great, thank you!"

"The next part (determinism/free will statements) will take about 10 minutes. I will present a few statements to you on my tablet. Please think deeply about each statement for exactly one minute. See if you agree with them. I will tell you when time is up!"

"Here comes the first statement (presented on screen)"

"Time's up! Here comes the nth statement!"

"Ok, those were all the statements! "

"Next up is the hard part, I hope you're not tired yet. After this part you will only have to fill in a small questionnaire though, so give it all you've got! I'm entering you into a little competition!"

"I will present difficult (multiple choice) questions on my tablet. Try to answer them correctly! Once you believe you know the answer to a question, tell me (by saying ready), and I will give you the correct answer (Pepper pronounces the answer). Please note down, on the form in front of you, if you got it correct (first measure of cheating)and if you understood the answer (second measure of cheating). Once you're ready for the next question, let me know! This is the current (fake, with an on average too high amount of questions to give an incentive to cheat) leaderboard, to give you an indication of what's possible. (amount of questions one can probably answer in those 10 mins) Try to answer as many questions as possible within the next 10 minutes (time on screen?)!

"Time's up! Please enter your total amount of answers and total amount of understood answers on my tablet!"

"Alright, as promised one more questionnaire and you're done!"

"(debriefing) That was it! The experiment is over. Thank you for participating, you may leave me now! Bye bye!"