

**Cues of being watched in the long run:**

*Evidence from two field experiments in public goods games*

[[1]](#footnote-1)

Master Thesis

Behavioral Economics

Erasmus University Rotterdam

September 2013

Royston P. Paulina

Student number: 272785

Supervised by Drs. Dennie van Dolder

Table of Contents

**Abstract3**

**Introduction4**

**Chapter 1: Literature Review6**

*§1.1 Classical repetitive public goods games* 6

*§1.2 Findings on eyes-like stimuli* 8

*§1.2.1 Findings on eyes-like stimuli in public goods games*8

*§1.2.1 Classification of findings on eyes-like stimuli in public goods games*10

*§1.2.2 Findings on eyes-like stimuli in dictator games*13

*§1.2.3 Findings on eyes-like stimuli in trust games*14

*§1.2.4 Concluding*15

**Chapter 2: Research Method**1**7**

*§2.1 Experiment 1* 17

*§2.2 Experiment 2* 18

*§2.3 Specifications of Experiment 1* 19

*§2.3.1 Specifications of Experiment 1: images*19

*§2.3.2 Specifications of Experiment 1: prices* 19

*§2.3.3 Specifications of Experiment 1: honesty box*19

*§2.3.4 Specifications of Experiment 1: subjects*19

*§2.4 Specifications of Experiment 2*20

*§2.4.1 Specifications of Experiment 2: common goods*20

*§2.4.2 Specifications of Experiment 2: prices*20

*§2.4.3 Specifications of Experiment 2: vending machine*21

*§2.4.4 Specifications of Experiment 2: communication*21

*§2.4.5 Specifications of Experiment 2: honesty box*22

*§2.4.6 Specifications of Experiment 2: subject pool*22

*§2.4.7 Specifications of Experiment 2: anonymity level*22

*§2.5 Hypotheses*24

**Chapter 3: Results26**

*§3.1 Results Experiment 1*26

*§3.2 Results Experiment 2*31

**Chapter 4: Discussion38**

**Literature41**

**Appendixes45**

*Appendix A: Classification of economic games*45

*Appendix B: Communication to the subjects of Experiment 2*47

*Appendix C: Communication to the Housing Committee regarding Experiment 2*47

*Appendix D: Communication to the subjects regarding endind of Experiment 2*48

*Appendix E: Pub trivia: application of a public goods game in a marketing campaign*48

*Figure 10: Data Experiment 1*49

*Figure 11: Data Experiment 2*49

*Figure 12: Cost overview Experiment 2*49

*Picture 1: Setup Experiment 1 (Treatment 1)*50

*Picture 2: Setup Experiment 1 (Treatment 2)*50

*Picture 3: Setup Experiment 1 (Change setup)*51

*Picture 4: Setup Experiment 1 (Treatment 3)*51

*Picture 5: Setup Experiment 2 (Treatment 1)*52

*Picture 6: Setup Experiment 2 (Treatment 2)*53

*Picture 7: Setup Experiment 2 (Treatment 3)*53

**Abstract**

Based on substantial amount of research on eyes-like stimuli to cue behavior, this paper researches whether these cues can be found in almost similar settings. And if this is the case, the question arises whether these effects prevail in the long run. Two field experiments have been conducted in two different settings in order to answer this research question. These experiments have been designed inspired by the experiment done by Bateson et al. (2006). The findings suggest that in neither experiment a significant effect of eyes-like stimuli could be found. However, several reasons are suggested in order to assess the implications of these findings.

**Introduction**

In a cleverly designed experiment by Bateson, Nettle, and Roberts (2006) they found that subject’s average contribution raised by almost three times when subtly cued by images of eyes. Their research was done in a real-world setting at the University of Newcastle at the Division of Psychology. In the public kitchen of this division the subjects (25 females and 23 males) could enjoy a cup of coffee and/or tea. Thereafter they could decide whether to contribute to the honor box or not. For years this system has been adopted and a price list for the coffee/tea/milk was provided in the kitchen. Moreover, each subject got a reminder mail twice a year regarding the contribution. No further control was in place, thus subjects were implicitly anonymous regarding their individual contribution. Bateson et al. (2006) used this setting by alternating, for ten weeks, the display of a banner on the price list containing either images of flower or a set of directly staring eyes. The measurement was done by calculating the weekly contribution to the honor box relative to liters of milk consumed. They found that the images of eyes increased the weekly contributions by 176% compared to images of flowers. Indicating that eyes-like stimuli can have a positive effect on social contributions in a public goods game conducted by means of a field experiment.

In a typical public goods experiment subjects, are initially endowed with an amount of money, tokens or another commodity[[2]](#footnote-2). Within a group of 4 to 10 subjects they must decide whether to make a contribution to the public good or not. Their individual contributions will be multiplied by a factor *k* that is greater than one but less than the number *n* of subjects. The contribution of an individual subject thus decreases by *k/n* for each unit of money but benefits the whole group by *k*. This game can be conducted as a one-shot game or a repetitive finite game. In order for subjects to maximize their individual pay-off, each one should contribute their entire endowment to the public good. Standard economic theory predicts that individuals will show opportunistic behavior, since they have an incentive to free ride on the contributions of other players. So the best strategy for each player (Nash equilibrium) is to contribute nothing to the public good if they face the same players in either a one-shot game or a repetitive one. This is known as the ‘strong-free-rider-hypothesis’. However, findings in public goods experiments contest this hypothesis, since subjects tend to make contributions to the public good. However, the cooperation levels tend to decline with repetition (e.g., Davis and Holt, 1993; Ledyard, 1995). Even in cases where subjects played a repetitive public goods game with three breaks in order to ‘restart’ the game, Cookson (2000) found a decline in the levels of cooperation.

Combining the abovementioned findings raises the question: will the positive effect of eyes-like stimuli prevail in the long run in a repetitive public goods game in a real-life-setting? It is debatable whether invoking subjects with a constant eyes-like stimulus in a real-life-setting of a repetitive public goods game, would lead to a constant rate of cooperation over time. Since, if eyes have a positive effect on contribution levels and the eyes-stimulus remains constantly present during the public goods game, it would follow that this positive effect could mitigate the decrease in cooperation over time. In this paper I will set out to answer the abovementioned question along with its relevant implications, by means of a literature review and experimental researches.

In the first chapter of this paper I will be conducting a literature review on the effects of eyes-like stimuli in the existent body of research. Mainly, I will be discussing previous public goods game experiments. The focus will be on four relevant characteristics which will be used to distinguish the findings of these experiments in order to apprehend the scope of the eyes-like stimuli in public goods games. Regarding the analysis of the effect on the long run of eyes-like stimuli; I will be conducting two field experiments using different subject pools and commodities, which will be described in the second chapter along with the statistical tests I will be performing. The third chapter will entail the results and the fourth the conclusions. In the final chapter I will be discussing the implications of the results and in the Appendixes you will encounter additional information.

**Chapter 1: Literature review**

In this chapter I will be reviewing the literature regarding classical[[3]](#footnote-3) repetitive public goods games, findings on eyes-like stimuli (positive findings, negative findings, public goods games, dictator games, and trust games), and a combination of these findings with the classical repetitive public goods games.

*§1.1 Classical repetitive public goods games*

As already mentioned, standard economic theory fails to explain the persisting tendency that people make voluntary contributions to public goods rather than free-riding on the contributions of others. This phenomenon has been a research topic since one of the earliest experiments on free-riding by Bohm (1972). Ledyard (1995) and Davis & Holt (1993) show, after extensive surveys of subsequent laboratory experiments on public goods games, that on average subjects contribute between 40-60% of their initial endowment to a public good. This contribution level typically tends to decline after repetition and eventually approach the free-riding level in the final repetitions.

In a study on framing effects in public goods games, Cookson (2000) found that framing a game with 32 repetitions as four games of 8 repetitions with small breaks in between of around 30 seconds, this caused a ‘restart-effect’. Furthermore, in this experiment he also offered different treatments regarding the payoff functions, and a comprehension task. One of his findings was that after each of the three breaks there was a substantial ‘restart-effect’ of increased contributions, averaging 22.8 percentage points per restart. Moreover, the speed of decay during each period increased with repetition: contributions decayed by 15.6, 21.9, and 30.4 percentage points respectively after each restart. However, the initial contribution levels of all the periods remained roughly constant at around 50%.

Combining this rate of decay and the initial contribution levels implies that the lower bound of the contribution levels obviously also declined in time throughout the 4 phases, however, still not reaching the zero-level of contribution as expected by standard economic theory. Several reasons have been researched for this cooperation to public goods and the decline of cooperation level over the past few decades which I have concisely summarized below:

**Repetitive classical public goods games**

**Human cooperative behavior among**

**Kin Non-Kin**

**Free-ride**

**Direct reciprocity Indirect reciprocity Tag-based cooperation**

**Intrinsic motivation Extrinsic motivation**

**Fairness Conditional cooperation Reputation**

**Strategies Learning**

In this paper, human cooperative behavior among non-kin will be discussed, since this is primarily relevant to public goods games. This social puzzle stems from the standard economic theory; why do subjects not entirely free-ride and still make non-zero contributions to public goods? From the Darwinian evolution theory a possible explanation might be at hand for the cooperation between non-kin related subjects: direct and indirect reciprocity. In the latter one the reward for an altruistic act is not expected from the recipient but from another member of the social group (Milinski et al., 2002). Another explanation is tag-based cooperation, offered by Riolo et al. (2001) where they found that cooperation can become established and sustained even without memory of previous actions of counter players. They found that not only do the agents not need continuing interactions with the same partners; moreover, the agents do not even need to observe the behavior of others or receive reports from third parties. ‘*Strategies of donating to others who have sufficiently similar heritable tags, even though such tags are initially arbitrary, can establish cooperation without reciprocity’* (Riolo et al., 2001, p.443). From the literature three main reasons can be found for indirect reciprocity; conditional cooperation, fairness preferences (altruism, ‘warm-glow’ of giving, inequity aversion, etc.) and reputation. Preferences for fairness can be viewed as an intrinsic motivation for cooperation, since these agents derive utility from contributing regardless of the actions of others (to some extent). It can be argued that reputational concerns can be labeled as extrinsic motivations, since these agents are prone to behave according to the social norms (e.g. contributing to a public good) in order to be positively assessed by others with possible positive future rewards to them, thus possibly using pro-social behavior as a signal. It can also be argued that conditional cooperation stems from an extrinsic motivation, since agents will cooperate if others do so. However, in a mixed group of extrinsically and intrinsically motivated agents, it could be argued that due to learning in repetitive interactions the intrinsic motivations could be crowded out, diminishing the average contributions. This would be the result of intrinsically motivated agents altering their strategies, however, still offering non-zero contributions at lower levels on the long run[[4]](#footnote-4).

Concluding, it can be stated that even though there is extensive evidence that the ‘strong free-rider hypothesis’ does not hold in public goods games, repetitions of the games lead to a decay in contribution levels as a result of several explanations. Moreover, even when everything else is held constant, restarting the games lead to relatively constant contribution levels in the first round of each phase. Thus, by incorporating these classical findings in findings on eyes-like stimuli could lead to a better understanding of the context and extent of positive effects on eyes-like stimuli in a repeated public goods game. The next paragraph will devote attention to these novel findings.

*§1.2 Findings on eyes-like stimuli*

In this paragraph I will be discussing the findings on eyes-like stimuli in public goods games in the literature, followed by a classification of these findings according to four main factors. In order to extend the scope of the effect of eyes-like stimuli, attention will also be devoted to a review of dictator games incorporating these stimuli in their research[[5]](#footnote-5). Also, some attention will be devoted to trust games incorporating these stimuli in their research. Concluding, the review on classical repetitive public goods games will be combined with the findings on eyes-like stimuli.

*§1.2.1 Findings on eyes-like stimuli in public goods games*

A growing body of research confirms that subtle social cues like images of eyes can have an effect on the decision-making process of subjects. In this paragraph I will be devoting attention to past studies concerning cues of being watched in public goods games.

As was mentioned in the introduction; Bateson et.al (2006) found a positive effect of eyes-like stimuli on the average contribution that was 2.76 times greater compared to the images of flowers. Bateson et al. (2006) used this setting by alternating, for ten weeks, the display of a banner on the price list containing either images of flowers or a set of directly staring eyes. The measurement was done by calculating the weekly contribution to the honor box relative to liters of milk consumed. They found that the images of eyes increased the weekly contributions by 176% compared to images of flowers. This indicated that eyes-like stimuli can have a positive effect on social contributions in a public goods game conducted by means of a field experiment.

In a laboratory[[6]](#footnote-6) public goods game experiment, Burnham & Hare (2007) found that incorporation of an animated robot with realistic eyes (Kismet), like ‘Clippy’ from Microsoft, increased the average contributions with 29%. In this experiment the animation was alternated every 0-2 seconds based on a random process in order to maximize the chance that at the moment of decision making, the neural activation in the eye-detection system of each subject increased.

In a field experiment by Ernest-Jones et.al (2011) they studied the littering behavior in a cafeteria on the campus of Newcastle University. Posters with either flower or eyes containing the message to clean one’s litter or an incongruent message were placed in this self-clearing cafeteria. These posters were randomly selected and put in place on each session during a period of 32 days[[7]](#footnote-7). They found a halving of the odds of littering in the presence of posters featuring the eyes, as compared to the posters featuring flowers. Moreover, increasing the party size increased the odds[[8]](#footnote-8) of leaving litter, whilst increasing the number of people in the cafeteria reduced it[[9]](#footnote-9). Thus, it was more likely for people to leave litter when the party size increased and less likely for people to leave litter when fewer people were around.

Findings of Ekström (2012) partially confirmed these effects of eyes-stimuli in another field experiment. He found that by placing pictures of eyes on recycling machines in a Swedish supermarket chain, subjects tended to give more to charity. Subjects had the possibility to either keep their cash after returning their cans and bottles or to donate this money to a charity by pressing adjacent buttons for either one option. When controlling for, store, day fixed effects, and using a proxy for store attendance, he found that pictures of eyes increased the donations by 30% during days when relatively few other people visited the store but no significant effect when the store was busy.

Powell et al. (2012) also found that displaying pictures of eyes in a supermarket in order to raise funds for charity had positive effects. On buckets for charity funds they displayed either a sticker of two eyes or three stickers of little stars. In both conditions the total length and width of the stickers were not more than 10cm respectively 3cm, indicating the subtlety of these adaptations. Their findings were in line with Ernest-Jones et al. (2011) and Ekström (2012); donations were on average 48% higher in the eyes-condition than in the stars-condition. Furthermore, they found that when the supermarket was quiet the eyes-effect increased the donations by 60%, whilst when the supermarket was busy this effect was 30%.

Concluding, it can be stated that from these findings it follows that eyes-like stimuli have a positive effect on pro-social behavior, in terms of contributions and exposing socially desirable behavior, in (real) public goods settings.

*§1.2.1 Classification of findings on eyes-like stimuli in public goods games*

In order to grasp the extent to which the abovementioned findings hold, this paragraph will devote attention to a more in-depth-analysis of these papers. This will be done by means of classifying the research in these papers based on 4[[10]](#footnote-10) main factors. The distinction made serves to grasp the context in which positive effects of eyes-like stimuli can be replicated in future studies.

As mentioned in the previous paragraph eyes-like stimuli have a positive effect on pro-social behavior. However, these effects were found after periodically altering the social cues, thus a *dynamic* approach (Bateson et al., 2006; Burnham & Hare, 2007; Ernest-Jones et al., 2011). Where Bateson et al. (2006) did these alternations on a weekly basis, Burnham & Hare (2007) on an interval of 0-2 seconds, and Ernest-Jones et al. (2011) in each session. However, it should also be mentioned that average charity contributions increased in real public goods settings even with the *static* use of the eyes-stimuli over a relatively long period of time. Ekström (2012) did this by altering the pictures once in the midst of a 12-days-period, while Powell et al. (2012) did so by leaving the charity boxes[[11]](#footnote-11) on the counter for a period of 11 weeks. Thus, the evidence suggests that it should not matter whether the static or the dynamic approach is used when researching eyes-like stimuli, both yield positive effects on pro-social behavior.

By examining the graphs of Ekström (p. 541, 2012) on the daily proportions of donors and daily proportions of donated amount, it remains ambiguous how this effect develops through time due to the peaks in the graphs in the weekends. Examining the graph of Powell et al. (p.1099, 2012) on mean donations (£) per thousand customers for eyes vs. control buckets in each week of the study, the suggestion may arise that the contribution levels of both treatments decline[[12]](#footnote-12) over time. This decline is in line with the findings in classical repetitive public goods games, as discussed earlier.

Thus, a priori it could be argued that a static approach of eyes-like stimuli will have a positive effect on pro-social behavior, but that this effect might decline over time.

The second classification that can be made between these papers is based on the *nature of the common good.* In the paper by Bateson et al (2006) subjects contributed to a public good which they could consume relative to the papers by: Burnham & Hare (2007); where subjects played a classic laboratory public goods game, Ernest-Jones et.al, (2011); where they measured littering behavior, Ekström (2012), and Powell et al., (2012); where they measured charity donations. The first two papers can be classified as offering a direct reward in terms of consumption goods and monetary pay-offs, whilst the last three papers as offering possible future rewards as a result of positive reputation-building. Specifically, in the last three papers by, Ernest-Jones et.al, (2011), Ekström (2012), and Powell et al., (2012), subjects did not receive a direct reward for their contributions. Thus, it is evident that when subjects are (unknowingly) participating in a public goods game, inducing them with eyes-like stimuli has a positive effect on their pro-social behavior no matter their reward for contribution. Possible explanations for this behavior can be found in the paper by Fehr & Fischbacher (2004). They found that people are often willing to punish others who break the social norms, which has been argued to be the major force maintaining pro-sociality in human societies (Gintis et al., 2003). According to Oda et al. (2011), eyes images have a positive effect on increasing pro-sociality but that the possibility of future rewards drives this behavior rather than the fear of punishment. Since their results suggested that their participants expected that their actions would enhance their reputation to a third party.

Concluding, it can be expected that even though the nature of the common good does not constitute a direct reward for contributing in a public goods game, subjects will make a higher than zero contribution (on average). This is due to the fact that reputational concerns and fear for punishment may play a role in the settings were subjects are being watched, or have the feeling of being watched.

The third classification is the *level of anonymity in public settings*. According to the findings of Ernest-Jones et al. (2011), Ekström (2012), Powell et al. (2012), eyes-stimuli had a greater effect when fewer people were around. Gächter and Fehr (1999), Andereoni and Petrie (2004), and Rege and Telle (2004) found that contributions in public goods games increased when there was no anonymity. Being observed and possibly evaluated by others while giving and also seeing what others give influenced their contribution levels. Van Dijk et al. (2002) already found that the perception of this evaluation is inherent to the forming of social ties between the subjects during public goods experiments. Masclet et al. (2003) stated that subjects increased their levels of contributions when the opportunity for other subjects to show their disapproval of their decisions was present.

Thus, it follows that the effects of eyes-like stimuli strongly depend on the level of (perceived) anonymity. Since, if real people are around, these real sets of eyes watching one’s behavior will diminish the effect of the experimental eyes-like stimuli. But when there are fewer people around, these experimental eyes-like stimuli will have a positive effect on pro-social behavior.

The last classification concerns the composition of the *subject pools* used in these researches. According to a survey done by Arnett (2008), of the top psychology journals 96% of the subjects were from Western industrialized countries which account for 12% of the world population[[13]](#footnote-13). As Henrich et al. (2010) stated; ‘*people from Western, educated, industrialized, rich and democratic (WEIRD) societies — and particularly American undergraduates — are some of the most psychologically unusual people on Earth*’. Researchers typically use students out of convenience and mistakenly stop to assess the implication of this method when generalizing their results. Gächter et al. (2010) found that cultural background has a substantial influence on cooperation between genetically unrelated individuals in otherwise identical environments. They also found that within cultures, cooperation was largely similar, but that there exist highly significant differences between cultures. Their research was based on the 10-rounds laboratory public goods game of Herrmann et al. (2008a) which used 16 subject pools from 6 different cultures as classified by Inglehart & Baker (2000). It should be noted however, that they even found this effect while conducting all their experiments with university undergraduates who were similar in age, shared an (upper) middle class background and usually did not know each other.

In an experiment conducted by Raihany and Bashary (2012) using Amazon Mechanical Turk, they also found culture to have a significant effect on donations[[14]](#footnote-14) in a cross-cultural one-shot dictator game[[15]](#footnote-15). Their cross-cultural subject pool consisted of 386 participants (145 females & 240 males), aged from 18 to 67 years old (mean of 27 years), self-reported annual income levels with great variance (less than$12.500 to over $100.000), and with educational levels varying from only high school up to postgraduate degree level. By using the Inglehart & Baker (2000) culture scheme they categorized their subjects (coming from 46 countries) into nine categories plus an additional category for those countries not assigned a world culture.

Using the culture-scheme of Inglehart & Baker (2000, p.35s & 36) the subject pools used in the papers on eyes-like stimuli in public goods games can all be classified as; *historically protestant* (situated in the upper right corner of the graph). Since the subject pools used in these papers were from Britain (Bateson et al., 2006; Ernest-Jones et al., 2011; Powell et al., 2012), the U.S. (Burnham & Hare, 2007) ,and Sweden (Ekström (2012). From Figure 4 in the paper by Inglehart & Baker (2000), these clusters are used to depict the coherence between interpersonal trust and GNP per capita. Again it can be seen that the cluster of historically protestant cultures score relatively higher (upper right corner of the graph) regarding ‘percentage who generally trust people’ and ‘GNP per capita’. Furthermore it is stated that this interpersonal trust is the basis for building the social structures on which democracy depends and for creating the complex social organizations. Thus it can be argued that the findings on public goods games could all be classified to one cluster, which appears to be the most social cluster. According to Inglehart & Baker (2000) a history of Protestant or Islamic or Orthodox or Confucian traditions, gives forms to cultural zones with distinctive value systems that persists after controlling for economic development. Moreover, economic development tends to push societies in a common direction, however, rather than converging in the same direction they find that these societies tend to move on parallel trajectories shaped by their cultural heritage. Thus, they argue that cultural heritage will prevail on the long run, deeming the relevance of these clusters.

Combining these facts yields another perspective on the positive findings of eyes-like stimuli in the already (extensively) discussed public goods games regarding the used subject pools. It can be argued that social norms, reputational concerns, altruism, reciprocity, and similar values might differ between cultures and hence on the effect of social cues. On the other hand, if all humans inherited the capacity to detect faces and eyes and react relatively more pro-social due to their neural architecture, the question remains what the overall effect would be of social cues in a cross-cultural setting.

*§1.2.2 Findings on eyes-like stimuli in dictator games*

In extending the scope of findings on eyes-like stimuli, I will devote this paragraph to the results in another type of game: the dictator game[[16]](#footnote-16).

In a laboratory dictator game Haley & Fessler (2005) found a positive effect on the average contribution. When comparing the eyes-treatment (where they displayed a stylized picture of eyes on the background of the computer screen) to the control treatment (where the eyes were replaced by text in the same color scheme), they found a positive effect on the average contribution of over 55%.

Rigdon et al. (2009) found in another laboratory dictator game that even three dots resembling[[17]](#footnote-17) the human face had a significant effect on male subjects. In the baseline 32% of the males gave $1 or more of their endowment and females 82%. In the eyes-condition this shifted to 79% of the males and 73% of the females. Their primary result is that the presence of a weak social cue impacts giving behavior of males but not females. After conducting a Logit analysis they found that male dictators are 3.35 times more likely than females to behave more altruistically in the presence of this weak social cue. Oda et al. (2011) also found that eyes have a positive effect on the average contributions in a laboratory dictator game of about 38% using the same stylized eyes like Haley & Fessler (2005).

Baillon, Selim and van Dolder (2013) did a research on the effects of social cues in interaction and individual choice tasks, where they used three conditions. One condition; with pictures of eyes of a statue directly gazing at the subjects, the second condition containing pictures of peer students without direct eye contact, and a control condition displaying pictures of rooms, hallways and buildings of the Erasmus University Rotterdam. Their subject pool was divided into four groups, each randomly participating in one of the four possible tasks.

The first interaction task was the so-called Joy of Destruction mini-game (Abbink & Herrmann, 2010). In the recent and growing literature the application of economic games on anti-social behavior, such as anti-social punishment of cooperators in public good settings (e.g. Gächter & Herrmann, 2009; Gächter et al, 2010; Herrmann et al, 2008), it was found that a considerable fraction of people are willing to pay to destroy the pay-off to their counterparty. The subject’s willingness to destroy markedly increased when their counterparty could not find out with certainty what caused the destruction.

In their control condition Baillon et al (2013) found that over 38% of the subjects destroyed their counterparty’s pay-off, whereas in the condition with eyes this was halved to over 17%. No significant difference could be found between the eyes-condition and the peers-condition (over 18% destruction). The second interaction task was a dictators game (Camerer, 2003), where they found that in their control condition subjects were, on average, willing to donate €9.75 of their initial €50 endowment (19.5%). In their eyes-condition this significantly increased to an average of €13.93 (27.86%), whilst the peers-condition did not significantly differ from the control condition (€9.11/ 18.22%).

Regarding the probability of donating, they found that in the eyes condition this was 73.63%, in the control condition 63.64%, and in the peers-condition 50.91%. The only significance they found in this respect was that subjects in the eye-condition were more likely to donate compared to subjects in the peers-condition.

The third task was an individual one where they employed a variant of the Ellsberg paradox (Ellsberg, 1961) regarding ambiguity aversion. Here they only found a significant effect in the peers-condition compared to the control-condition, and not for the eyes-condition. In the last individual task on the choice between simple and compound gambles (Bar-Hillel, 1973) again there was no significant effect of the eyes-condition relative to the other two conditions. Only in the peers-condition there was a marginal significant effect on the choices made compared to the other two conditions.

Concluding, it can be stated that the positive effect of eyes-like stimuli can also be found in other settings rather than only public goods games.

*§1.2.3 Findings on eyes-like stimuli in trust games*

In a meta-analysis conducted by Johnson and Mislin (2011) they could not find conclusive evidence on the anonymity in trust games. This fact strokes with the idea that in trust[[18]](#footnote-18) games the Responder has influence on the action of the Proposer by either rejecting or accepting the offer which has consequences for the payoff of the Proposer. Reciprocity and altruism are the most influential in this decision and thus incorporating eyes-like stimuli does not yield any effect on the average contributions of the Proposers and/or the Responder (Cookson, 2010). For this reason this game has been omitted from the analysis in this paper.

*§1.2.4 Concluding*

From this literature review it can be concluded that subjects playing a public goods game will, on average, make a positive contribution to the public good. However, this effect diminishes when the games are repeated and even after restart. However, the levels of contribution tend to rise to a relatively constant value of 50% of the initial endowment in the early rounds after each restart. It has also been shown that inducing subjects with eyes-like stimuli has a positive effect on the contribution levels to public goods games.

Moreover, the presentation of these stimuli merely needs to resemble the cues of being watched by real eyes. Dots in a v shape, robot-eyes, two-dimensional eyes on a piece of paper, stickers of eyes, and direct gazes from statues on a website, are all effective in enhancing pro-social behavior. Moreover, this effect is not only limited to public goods games, but also evident in dictator games. Due to the nature of trust games, no evidence has yet been found on this eyes-effect.

Several factors can be alluded to in the literature as possibly influencing the effects of eyes-like stimuli. Examples of these factors are: level of perceived anonymity, nature of the common good, composition of the subject pool, and the static versus dynamic use of eyes-like stimuli.

In this paper I have made a first attempt of using these factors. I tried to do so by designing two field experiments on repetitive public goods games where the level of perceived anonymity was relatively high, the nature of the common good constituted a direct reward, the composition of the subject pool was relatively more diverse compared to previous researches on eyes-like stimuli in public goods games, and the eyes-like stimuli were static.

Derived from the literature review it can a priori be reasoned that these experiments will show a positive effect on the average contributions. Moreover, a decline of the levels of contribution is possibly to be expected in the long run. Furthermore, this decline might not only be the result of the reasons mentioned earlier (reciprocity, motivation, etc.), but also from habituation[[19]](#footnote-19). This reasoning is derived from the dual-process theory of response habituation (Groves & Thompson, 1970; Rankin et al., 2009). They state that repeated application of a stimulus might first show an increase in responsiveness followed by a progressive decrease in some parameters of a response to an asymptotic level (in this case that could be the y-axis on the level of contribution).

In the next chapter I will continue with an elaboration on my research method in order to test the positive effects of eye-like stimulus and the development of contribution levels over time in repeated public goods games.

**Chapter 2: Research Method**

As previously mentioned, the research method used in this paper will be two field experiments in a public goods game setting based on the experiment by Bateson et al. (2006). The outlines of the experiments will be discussed followed by their specifications.

*§2.1 Experiment 1*

As was the case with the experiment conducted by Bateson et al. (2006), a setting was used for this experiment where the provision of ‘free coffee and tea’ in combination with an honesty box was already common. This field experiment was conducted at Soundport Rotterdam[[20]](#footnote-20), which is a four-floor complex housing four music studios on each floor for (semi-)professional and amateur musicians. Specifically, this experiment was conducted only on the fourth floor of this complex. The practical reason behind this was that only musicians on this floor were familiar with the concept of the honesty box. The other studios on the lower floors all had their own provision of coffee and tea in their studios.

For the period of 9 weeks an A5-sized notice was displayed on top of the counter where the equipment for making tea and coffee had been placed. This notice contained information regarding the prices of tea, milk and coffee along with a banner with an image of flowers or eyes.

The main difference with the original experiment by Bateson et al. (2006) was that instead of altering between flower- and eyes-images on a weekly basis, this was only done twice in this experiment. During the first four weeks, a banner of a flower-image was showed on top of the price list which was kept the same for this whole period. During the following four weeks, this banner was replaced by an image of eyes which was also kept the same for this whole period. In the last week (9th week) this eyes-image was replaced by another eyes-image to control for the effect of a different pair of eyes on the average contribution.

The reason for starting with images of flowers was to convey a more neutral starting point for the experiment. Moreover, it could be argued that the possible increase in responsiveness of a stimulus followed by habituation would in this case be easier to distinguish, since it is expected that the effect of eyes-images would be larger than those of flowers. Additionally, it should also be mentioned that previously no price list was used in this setting, thus by starting with the flower-images the subjects would get accustomed to this new setting.

Hence, by using the approach outlined above, it can be tested whether the stimuli of the eyes will prevail in the long run. For an illustration please see Pictures 1 & 2 & 4 in the Appendix.

*§2.2 Experiment 2*

This experiment was conducted in a building for student housing and young professionals (55 subjects). Since 2010 a contract has been signed with a company which delivered a machine for candies, snacks and cold beverages. However, the reliability of this machine issuing products after payments has been low for over a year now, despite several attempts to address this issue. Since then a notice has been posted on this machine stating that the use is at own risk.

In this experiment the lack of reliability of the machine was used as the main argumentation for the setup. In order to provide the subjects with the service that was lacking from the vending-machine-company, a basket with products was provided from which they could freely pick the desired product(s) and then decided whether and how much they would contribute to the honesty box. Obviously these subjects receive guests which may not feel obliged to contribute as they might perceive it as a one-shot-game with no repercussions. In order to trying to control for this side-effect, the basket and honesty box were placed in the laundry facility which is only accessible with a key that only subjects living in the building possess.

Regarding the period, duration, choice of images, alteration of the images, and the intuition behind these choices, this experiment was exactly similar to Experiment 1. However, in order to mimic the vending machine, two price lists were placed instead of one. Thus, one price list was hung on the wall on eye-level and one beside the basket with goods. For an illustration please see Pictures 5 & 6 & 7 in the Appendix.

The duration of both these experiments was from Monday 29th of April until Monday 1st of July 2013. This covered a period of 9 weeks where the last week was for control purposes. The intuition behind using two blocks of 4 weeks was to control for possible income-level-/ week-effects due to the fluctuations of income during the month. Students with a Dutch scholarship received their pay on 24th of April, 24th of June and 24th of May[[21]](#footnote-21), where only the last payday coincided with the same day the change of eyes-image in week 9. The other paydays all fell one week before weeks 1 & 5 of the experiments. According to Dutch Labor Laws[[22]](#footnote-22) every worker should get their salary at least once a month with the actual date specified by the employer. Thus, the payday of the young professionals could have varied within the month.

Regarding the musicians it is customary that they receive their payment after a performance at a random day in the same month or any following month. Since this information was difficult to retrieve while conducting a field experiment, this was solved using two blocks of four weeks. By doing so it could be argued that possible confounding factor related to availability of money throughout the weeks has been mitigated.

*§2.3 Specifications of Experiment 1*

I will now continue with the specifications of Experiment 1. Please note that the use and choice of flower- and eyes-images were the same for both experiments. Also both experiments were conducted simultaneously.

*§2.3.1 Specifications of Experiment 1: images*

With respect to the images, these were also black and white and the same size as in the original experiment (150x35mm) by Bateson et al. (2006), however; only two of the original 10 photos were used. Specifically, the eyes- and flower-images which had relatively the highest and respectively the lowest effect on the contributions in the original experiment were used. The argumentation behind this was that, there is a difference between the images of the eyes and flowers related to the contributions measured as can be seen in the graph presented in the paper by Bateson et al. (2006). However, possibly another explanation for the relatively larger effect of the first set of eyes-image can be due to the fact that it was the first alternation in the familiar setting. It should be noted however, that it may seem that male eyes and/or the angle of the face have a greater effect on the average contributions. Obviously this cannot be properly analyzed since other confounding factors might play a role, e.g. amount of subjects in that week. Following these arguments I will use the eyes-images of weeks 1 & 9, and for the flower-images that of week 6 (Bateson et al, 2006).

*§2.3.2 Specifications of Experiment 1: prices*

Regarding the prices; average price of these commodities in other public places[[23]](#footnote-23) were used. The prices were listed in the order of 30, 50 and 10 eurocents for tea, coffee and milk respectively.

*§2.3.3 Specifications of Experiment 1: honesty box*

The honesty box used was transparent in this experiment for the practical reason that this has been the case already for years in this setting. Thus, in order to try to maintain the original setting at the music studios, this was not to be replaced by another one. In accordance with the building manager, this box was emptied weekly between 12:00-14:00 hours[[24]](#footnote-24).

*§2.3.4 Specifications of Experiment 1: subjects*

Regarding the amount of subjects, unknowingly, participating in this experiment, only an educated guess is possible. This is due to the fact that there are four studios on the fourth floor which are rented out to different music bands and/or individual musicians. A preliminary interview with the building supervisor, who happens to be also responsible for the provision of coffee/tea/milk and entitled to the earnings in the honesty box, yielded an estimate of around 50 musicians per week visiting the fourth floor. However, it was not obvious how many of these subjects would make use of the facilitation of hot beverages. In order to circumvent this problem, an almost similar procedure to that of Bateson et al. (2006) was followed (regarding the measurements of the consumption relative to the monetary contributions). Almost similar, for the practical reason that it is not a custom in the Netherlands to drink tea with milk as is the case in England (Newcastle in the original experiment by Bateson et al., 2006). Moreover, measuring only liters of milk used whereas it was also possible to use measurements of coffee and tea would have led to an unnecessary loss of data points. Thus, measuring of the amounts used was done by deducting the amount of tea bags used, grams[[25]](#footnote-25) of milk consumed, cubes of sugar consumed, and grams of coffee used for the machine[[26]](#footnote-26).

With respect to the communication regarding the implementation of the price list, as was the case in the original experiment, there will be none.

*§2.4. Specifications of Experiment 2*

I will now continue with the specifications of Experiment 2.

*§2.4.1 Specifications of Experiment 2: common goods*

The basket of goods contained cans of Coca-Cola and Fanta (both 33cl), Croky Chips Bolognese (100 grams), Croky Chips Naturel (100 grams), Mars (47 grams) and Snickers (50 grams) in the respective ratio of 4:2:2:2:4:4. These products and ratios were derived from the current assortment of the vending machine, which in turn was derived from the relative amounts sold when filling up the assortment weekly by employees of the vending-machine-company. Regarding the nature of the products used, e.g. unhealthy, it could be argued that the basket should have also contained healthy products. However, the ratio of healthy versus unhealthy products in the current assortment of the vending machine is negligible. Moreover, mixing these unhealthy products with healthy products could lead to confounding factors regarding the findings, thus only unhealthy products have been used.

*§2.4.2 Specifications of Experiment 2: prices*

The prices of the goods were all set at almost the same level as the prices of the same products in the vending machine, namely €0.80 for all products. The vending machine prices are €0.80 for the beverages and chips, and €0.75 for the chocolate bars. For practical reasons all prices were set at €0.80[[27]](#footnote-27).

*§2.4.3 Specifications of Experiment 2: vending machine*

In order to control for the variance in reliability of the vending machine[[28]](#footnote-28) and the possible confounding factor that only subjects that were not willing to pay would have made use of this arranged basket of goods, the machine was completely placed out of order by unplugging it from the electricity socket. Moreover, this theory was tested by extending the duration of this experiment with one week, thus making it ten weeks. This was done by maintaining the setup of week 9 but with the difference that the machine was plugged back in at the start of week 10[[29]](#footnote-29). It was found that the weekly average contributions declined from €0.55 to a mere €0.14, thus supporting this reasoning. No other changes were made to the setting except that subjects received communication that the vending machine was working again at the end of the ninth week and the machine was plugged back in. This was no deception, since during the experiments the company noticed the unplugged machine and fixed it in the eighth week of this experiment and the machine was switched back on at the end of week 9.

*§2.4.4 Specifications of Experiment 2: communication*

The communication[[30]](#footnote-30) about the basket of goods was done by e-mail, posting in a closed Facebook-group[[31]](#footnote-31), posting on our website[[32]](#footnote-32), sticking a poster on the wall of the elevator, and sticking a poster on the notification wall-board in the hallway of the main entrance[[33]](#footnote-33). This was all done on the 28th of April 2013 and stayed for the duration of this experiment, except for the mailings, Facebook- and website-postings which were sent only once. In these notifications subjects were attended on the fact that the vending machine was totally out of order, but that the alternative was to go to the washing facility to grasp a snack and/or beverage. Subjects were kindly but firmly requested to make a contribution in the honesty box upon taking one or more products. It was also mentioned that this basket would be refilled daily. No implicit or explicit mentioning was done about any repercussions if one did not contribute after taking one or more products, neither did I mention that this basket of goods would have been available only for a restricted period. Please see Appendix B, C & D for the communication notes.

Doing this experiment in a confined area where only[[34]](#footnote-34) residents of the building had access to would limit the possibility that strangers visiting the building would see these posts hanging around, locate the basket and decide to take one or more products without making a contribution, since in that case it could be perceived as a one-shot game without repercussions. The basket of goods was placed in the washing facility which is on the first floor of the building in contrast with the vending machine that is placed in the hallway of the main-entrance. The placement of the basket of goods on the first floor could be perceived as similar to that of vending machine in the hallway, since in both cases subjects had to exert relatively low effort in order to get a product. This is due to the architecture of the building which has it that there are no apartments situated on the same level as the main-entrance; every resident should either take the staircase or the elevator in order to get to their apartment.

*§2.4.5 Specifications of Experiment 2: honesty box*

As Soetevent (2005) found that donations in churches increased when the collection bags were open relative to close, in this experiment the honesty box was transparent in order to avoid confounding factors related to the transparency and willingness to pay. Another finding of Soetevent (2005) was that the initial contributions in the open bags were used as a benchmark by others. Using this reasoning, by default the honesty box always contained a minimum of €1.60 even after emptying. This amount was divided into eight 5 eurocents, two 10 eurocents, and five 20 eurocents. It could be argued that subjects would not really perceive this to be the full contribution of two products, although, the amount of coins resembles that contribution is the norm.

Measurement of the average contributions was done by dividing the contributions made in the honesty box by the amount of products taken on a daily basis. The basket was refilled every day between 12:00-14:00 hours to the ratio earlier mentioned. Furthermore, it should be mentioned that double-sided tape was stuck on the inner top of the honesty box[[35]](#footnote-35). This measure was taken in order to check whether subjects would try to take money out of the honesty box by turning it upside down. As Ariely (2012) argued; people would not mind stealing a stack of printing paper, but not €3.50 out of the office petty cash box. He argues that people have an inner moral that can justify the stealing of relatively cheap goods, but are averse from stealing the equivalent monetary value of these goods.

*§2.4.6 Specifications of Experiment 2: subject pool*

As was found by Henrich et al. (2010), Arnett (2008), Gächter et al. (2010), Herrmann et al. (2008a), Inglehart & Baker (2000), Raihany and Bashary (2012), cultural background may have an effect on the findings in experiments. The subject pool in this experiment consisted of 55 subjects (22 males & 33 females), consisting of 20 full-time students, 17 full-time working young professionals, 18 full-time students with part-time jobs. Unfortunately no information could be gathered regarding the income levels of the subjects, since they have not been debriefed about the experiment and thus giving the opportunity for a questionnaire. Regarding the ethnicity of these subjects the division was as follow: 19 from the Netherlands, 10 from Curaçao, 5 from Aruba, 5 from Surinam, 3 from Indonesia, 3 from China, 2 from Morocco, 2 from Turkey, 2 from Russia, 1 from Brazil, 1 from Serbia, 1 from Nigeria, 1 from Ukraine, and 1 from Iran. Applying this diverse division of ethnicity to Inglehart & Baker’s Figure 3, page 35 (2000) on the persistence of traditional values clustered by cultural zones; it can be argued that every cultural zone is represented. This is after pooling subjects from Curacao, Aruba and Surinam in the ‘Historically Catholic’-cluster and the subject from Serbia into the ‘Historically Communist’-cluster based on their geographical location.

The average age was 26.19 with a standard deviation of 2.73.

It should be noted that of these 55 subjects 5 never used the washing facilities in the building[[36]](#footnote-36), however, they still had access to this room, and therefore they have not been excluded from this pool.

*§2.4.7 Specifications of Experiment 2: anonymity level*

In order to capture the behavior of all 55 subjects, strict secrecy was implied. However, since I am only the building manager and do not constitute the whole management by myself, fellow management-members (Housing Committee) were not informed about this experiment. Please take dully notice that I did not resort to deception of my colleagues. During the years they have come to know me[[37]](#footnote-37) and acknowledge my resentment towards my own initial proposition of the use of a vending machine. They also know that I have made several fruitless attempts to inform the company that their service is lacking. The way I went around this issue was by stating that I wanted to try something new and that if I could break even I would be content. In case I made any profit, the setup of this project would be taken over by the building association after I was done with my project. Please see Appendix C for the communication note to the Housing Committee.

Although it was not possible to follow individual contributions due to the nature of this experiment, an online application proofed insightful regarding which subjects visited the laundry facility at what time and on what day. This was derived from the online washing-reservation-schedule which use is compulsory in order to use the machines. Every subject could access this website and log in using their private username and password and proceed to make one or more reservations. On average most subjects made reservations between Friday and Sunday or in the evenings due to their busy time-schedules. Coincidence[[38]](#footnote-38) has it that a new rule had been approved by the management and the residents stating that using the facilities without a valid reservation could result in an immediate fine of €25. This rule has been implemented starting from May 1st of 2013 for an undefined period. Thus, the propensity that all subjects would be inclined to make a reservation in order to use the facilities would at least stay the same or increase. However, this fact still did not control for the possibility that subjects would go to the facilities not to wash but for the basket of goods.

Using this reservations-website helped in order to assess the social multiplier effect (Ekström, 2011), as reservations that overlap regarding the starting time and ending time for two different subjects could be tested for statistical significance on the average contributions. However, the washing programs ranged from 35 minutes up to 56 minutes, whereas residents could make a reservation for a maximum of 90 minutes. So, it was still not possible to conclude that, if the reservations overlapped, two subjects had been in the laundry facility at the same time.

Moreover, another washing rule stated that each resident has 10 minutes of extra time to collect his/her clothes before the other resident is entitled to remove these from the machines. In practice[[39]](#footnote-39) this results in most residents going to the laundry facility ten minutes after their reservations, and thus minimizing the chances of having to wait for another resident. Moreover, I limited my presence in the laundry facility by reducing my washing reservations and trying not to overlap with other residents’ reservations. Moreover, before daily gathering the data, I would first check the reservations-schedule in order not to meet with any subjects doing their laundries. I remained successful in these efforts.

*§2.5 Hypotheses*

I will now continue with the hypotheses that were formulated for Experiment 1 & 2. The test results of these hypotheses can be found in the next chapter.

As was found in fMRI studies by Emery (2000) and Haxby et al. (2000), humans have a dedicated neural architecture which upon the detection of faces and eyes, activates automatically and uncontrollably even if these faces and eyes were represented two-dimensionally. In studies with chimpanzees Hare et al. (2001 & 2006) found that subordinate individuals refrained from taking prized food when they were in someone’s view depending whether they could see another individual’s face and eyes. Thus, changing from four weeks of flower-image to eyes-image in week 5 would induce the stimuli of being watched which in turn will affect the average contribution positively. If there is a positive effect of eyes-like stimuli, this would result in the average contributions in treatment 2 (eyes) to be significantly higher than the average contributions in treatment 1 (flowers). This was tested using the following hypotheses:

*H1.1: average weekly contributions from treatment 2 > average weekly contributions from treatment 1 (Mann-Whitney U / Wilcoxon Rank Sum test) for Experiment 1*

*H1.2: average weekly contributions from treatment 2 > average weekly contributions from treatment 1 (Mann-Whitney U / Wilcoxon Rank Sum test) for Experiment 2*

As was mentioned before; the dual-process theory of response habituation (Groves & Thompson, 1970; Rankin et al., 2009) states that repeated application of a stimulus might first show an increase in responsiveness followed by a progressive decrease in some parameters of a response to an asymptotic level. Combined with the findings on classical repetitive repeated public goods games that the average contributions tend to decline over time, it can be argued that this would also be the case in both experiments under both treatments. Following these argument I defined my next hypotheses as follows:

*H2.1: The average contributions will diminish during weeks 1 to 4 and during weeks 5 to 8 (analyzing the trend in the graph of the average contributions on the weeks) in Experiment 1*

*H2.2: The average contributions will diminish during weeks 1 to 4 and during weeks 5 to 8 (analyzing the trend in the graph of the average contributions on the weeks) in Experiment 2*

At the end of week 8 the eyes-image in both experiments were replaced by another image of eyes at the start of week 9. If the dedicated neural architecture of the subjects automatically activates upon the detection of a ‘new’ set of eyes, it could be argued that this could trigger a relatively higher average weekly contribution in week 9 compared to week 8. This reasoning is also in line with the dual-process theory of response habituation (Groves & Thompson, 1970; Rankin et al., 2009). Following these argument I defined my next hypotheses as follows:

*H3.1: The average weekly contribution of week 9 will be higher than that of week 8 in Experiment 1*

*H3.2: The average weekly contribution of week 9 will be higher than that of week 8 in Experiment 2*

**Chapter 3: Results**

*§3.1 Results Experiment 1*

As was mentioned earlier, the weekly average contribution was calculated as the amount of money weekly contributed divided by the amount of grams of coffee, sugar cubes, grams of milk, and tea bags weekly consumed. These measurements yielded 4 independent observations for treatment 1 (assuming each decision about, and the amount of the contributions were made independently) and 4 independent observations for treatment 2. As can be seen in Figure 10 in the Appendix, the average contributions as a function of milk, sugar and tea consumed had great variability. Average contribution as a function of milk ranged from €0.01 to €6, this was possibly due to the fact that the subjects did not regularly drink their coffee or tea with milk. For sugar and tea the same reasoning could be applied. Moreover, the consumption in absolute values was not suited for analysis since too much weight was put on these average contributions due to the limited use of these goods. For example; in week 3 only one teabag was used. I omitted this problem by using the average contribution as a function of the amount of grams of coffee consumed, since the data was more suitable for this analysis.

However, despite the enthusiasm and willingness to cooperate in this experiment, the building manager of this facility decided to change the setup amidst week 8 of the experiment. This was done due to the accommodation of an additional kitchen appliance, but possibly also due to the discomfort he felt when entering the kitchen during the weeks with eyes-images. In a previous week he mentioned that he started feeling discomfort as a result of the eyes-images. The alteration of the experimental setup was not discussed with me, so I could not mitigate this change. This resulted in the omission of the data in weeks 8 and 9. For an illustration of the change in the experimental setup please see Picture 3 in the Appendix. Since the income of musicians varies quite a bit during the months, I deemed it not necessary to omit the data from week 4 in order to control for possible income-level effect.

From Table 1 it can be seen that the average contribution as a function of grams of coffee consumed under treatment 1 was €0.1093657 with a standard deviation of 0.0381347 and under treatment 2; €0.1577183 with a standard deviation of 0.1142357. From Table 2 it can be seen that the overall average contribution as a function of coffee grams consumed was €0.130083.

Table 1



Table 2



In order to test for the normality of the distribution, a Shapiro-Wilks test was conducted. As can be seen in Table 3, this yielded a p-value of 0.290 and a z-score of 0.553, concluding that the distribution of sample population was not normal, as a result of the rejection of the null hypothesis stating that the population sample follows a normal distribution. This can also be seen in Figure 1 depicting a histogram on the frequency.

Table 3 Figure 1



For Experiment 1 the distribution of the weekly averages of contributions in the two experimental conditions (flowers versus eyes) was compared using a non-parametric statistical test, due to the violation of the first assumption of parametric tests. Specifically, the Mann-Whitney U test/Wilcoxon Rank Sum test was used, since two independent samples were to be tested whether they came from the same population.

*H1.1: average weekly contributions from treatment 2 > average weekly contributions from treatment 1 (Mann-Whitney U / Wilcoxon Rank Sum test) for Experiment 1*

From Table 4 it can be seen that there is no significant effect between treatment 1 and 2, z-score is -0.354 with a p-value of 0.7237. Even after omitting the data point of week 4, this yields no significant effect with a z-score of -0.218 with a p-value of 0.8273, as can be seen in Table 5. Concluding that; the null hypothesis that the difference between the medians arose through sampling effects cannot be rejected. Thus, eyes-images had no significant effect on average contributions as a function of coffee grams consumed in Experiment 1.

Table 4 Table 5

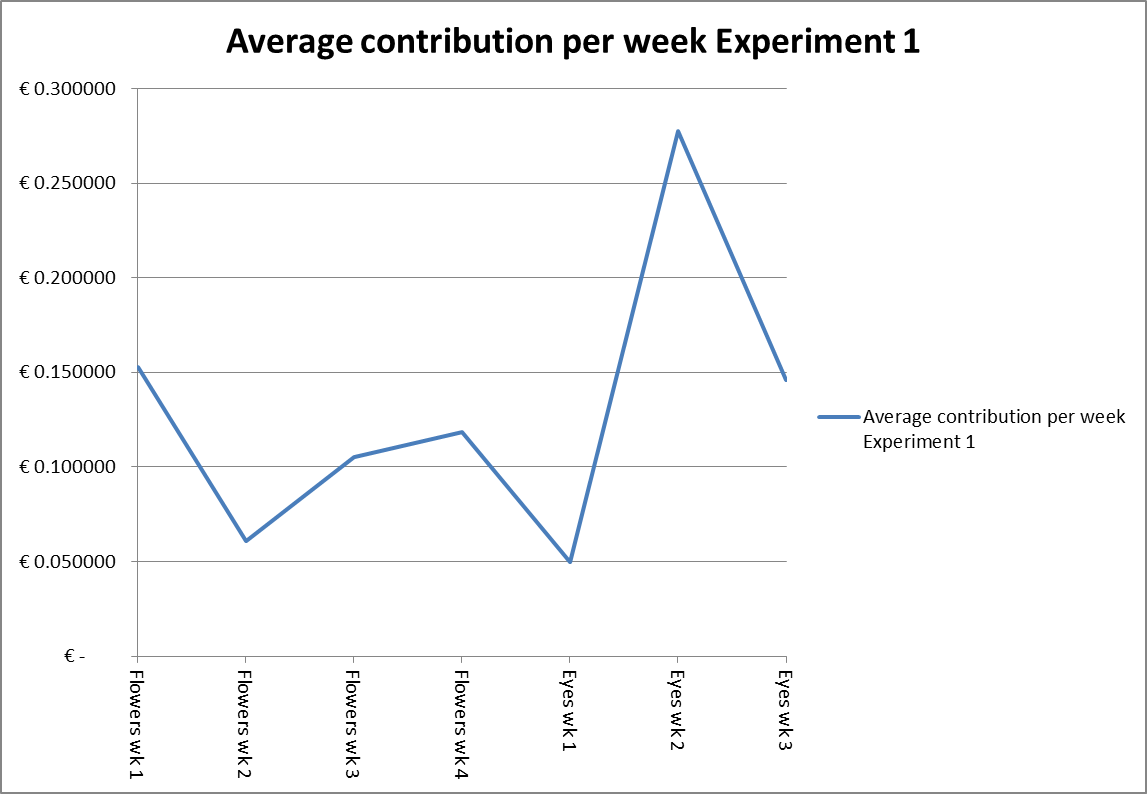
 

*H2.1: The average contributions will diminish during weeks 1 to 4 and during weeks 5 to 7 in Experiment 1*

By analyzing Figure 2 it can be seen that the average weekly contributions as a function of coffee grams consumed decreased from week 1 to 2 but then increased in weeks 3 and 4 of treatment 1. Several explanations might be found for these phenomena; e.g. subjects could have had relatively less income to spend in week 2, or a large group of subjects used the facility at the same time while the first subject did not pay and induced the social multiplier effect of not paying. However, since no actual observations of individual or group contributions were monitored, there could have been also a combination of these reasons throughout the days of the weeks. Please note that despite the fact that the contributions were decreasing or increasing, the overall average contributions for weeks 1 to 7 were only 26% of maximum contribution for coffee of €0.50.

A relatively large drop can be observed in the average contributions from week 4 to week 5, this might be due to the fact that more people felt uncomfortable by the eyes-image and consciously ignored the price list and consecutively made zero contributions to the public goods. It should also be noted that in weeks 4 and 5 there was no provision of coffee milk. This could also have triggered a negative attitude towards contributions if subjects that are used to drink their coffee with milk were restrained in their utility.

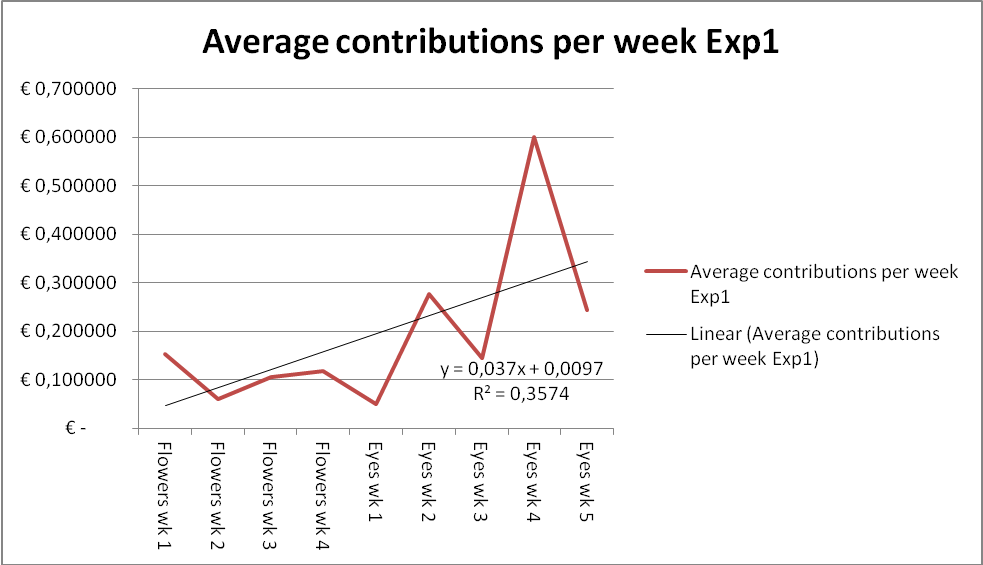
Figure 2



*H3.1: The average weekly contribution of week 9 will be higher than that of week 8 (analyzing the graph of the average weekly contributions for weeks 8 & 9). Experiment 1*

Obviously this hypothesis could not be tested due the change of the experimental setup. A comparison of the contributions levels of week 8 and 9 was not possible due to the fact that the change of the setup occurred in the middle of week 8 and the data on week 9 was for a full week. However, measurements have still been done for weeks 8 and 9 despite this fact. It can be argued that due to the change of the setup the average contributions increased to a level €0.10 beyond the price for coffee of €0.50 (Figure 3). However, this might prove to be speculations, since the experimental design had been compromised.

Figure 3



In order to test ex post what the optimal sample size should have been, a power test calculation was performed using the current data with a one-sided alpha of 5%. From Table 6 it can be observed that for both treatments the optimal sample sizes were 54. This coincides with the estimation of the building manager about the amount of musicians visiting the studios on a weekly basis. From Table 7 it can be observed that the power of this test was 0.9911 which is quite high, thus 99.11 percent of the null hypotheses were correctly rejected. For the calculation of the power test the current data and an alpha of 5% were used again.

Table 6 Table 7

* *

*§3.2 Results Experiment 2*

I will now continue discussing the results of Experiment 2. As was mentioned earlier, the daily average contirbution was calculated as the amount of money contributed divided by the amount of items taken from the basket of public goods. These measurements yielded 20 independent observations for treatment 1 (assuming each decision about, and the amount of the contributions were made independently) and 23 independent observations for treatment 2. This was after omitting 13 observations due to the fact that on these days no items were taken and obviously no contributions were made. The argument behind this reasoning is the fact that by incorporating zero-points from non-actions, could contaminate the analysis including the zero-points where subjects took items without making any contributions. Where there to be relatively more observations in treatment 1 compared to treatment 2 after this ommission, it could have been argued that potentially some of this effect of non-action could have been attributed to e.g. a fear-induced reaction by subjects due to the eyes-images. However, with 20 independent observations in treatment 1 and 23 in treatment 2 after omission, it can be argued that this possibly was not the case in this experiment.

In order to test whether the distribution of the daily average contributions was normally distributed, a Shapiro-Wilk test was conducted. With a p-value of 0.00459 the null hypothesis was rejected that the data was from a normally distributed population. Concluding that the data in this experiment was not normally distributed, resulting in the use of non-parametric statistcal tests. The results of this test can be viewed in Table 8.

Table 8



Assessing the normality of the data graphically by the use of histograms confirmed this finding as can be seen in Figure 4, showing the distribution for the whole data set. In Figure 5 the moderate skewed distribution of the data in treatment 1 is shown and in Figure 6 the uniform distribution of the data in treatment 2.

Figure 4 Figure 5 Figure 6



In order to test whether the two distributions were drawn from the same population, the non-parmetric two-sample Kolmogorov-Smirnov test for equality of distribution was conducted. From Table 9 it can be seen that the p-values were not significant on a confidence interval of 95%, implying that the null hypothesis was accepted stating that the two sample distributions were not significantly different from each other.

Table 9



From Table 10 it can be seen that the average contribution for treatment 1 was slightly lower than that of treatment 2 in Experiment 2; namely €0.4865 compared to €0.4878261 with standard deviatons of respectively 0.3428407 and 0.3464227. Table 11 shows the overall average contribution of €0.4872093 which is 60.9% of the overal average value of the public goods.

Table 10 Table 11



From Table 10 it can also be seen that the variances in both treatments are not identical (Treatment 1; 0.1175397, Treatment 2; 0.1200087). In order to assess the equality of variances, a non-parametric Levene’s Test was conducted, since if the distribution of the sampling population is skewed, the non-parametric test has more statistical power (Nordstokke & Zumbo, 2010). The p-value was found to be 0.847, which is statistically insignificant, thus accepting the null hypothesis that the two samples have equal variances. The results can be viewed below in Table 12:

**Table 12**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | |
| Variance (price/item) | | | | | |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 1,373 | 1 | 1,373 | ,038 | ,847 |
| Within Groups | 1483,833 | 41 | 36,191 |  |  |
| Total | 1485,206 | 42 |  |  |  |

Since the distribution of the data was not normal due to many full or 0 payments and the two sample sizes were unequal, I used non-parametric tests. For Experiment 2 I compared the distribution of the daily averages of contributions in the two experimental conditions (flowers versus eyes) using statistical tests suitable for these data. For the first hypothesis the Mann-Whitney U test/Wilcoxon Rank Sum test was used, since two independent samples were to be tested whether they came from the same population.

*H1.2: average weekly contributions from treatment 2 > average weekly contributions from treatment 1 (Mann-Whitney U / Wilcoxon Rank Sum test) for Experiment 2*

As can be seen in Table 13, the z-score is 0.492 which yields a p-value of 0.6230. This p-value is insignificant at the 95% confidence level. It can thus be concluded that the median of the average daily contributions for treatment 1 and 2 are not significantly different from each other, yielding no effects of eyes-images on average contribution levels in this experiment[[40]](#footnote-40).

Table 13



There was potentially circumstantial evidence that on some Wednesdays items were taken without contribution by one or more (external) cleaners of the building. The teams varied across the weeks so no trend on this behavior was observable in the data. However, even after omitting data points on every Wednesday from the dataset, this yielded no significant result for the eyes-images. The p-value was in this case 0.6893 with a z-score of 0.4000. Data points on Wednesdays have been included in further analysis due to this negligible difference.

Moreover, as can be seen in Table 14; the difference between week 4 and week 5 reported a z-score of -1.324 and a non-significant p-value of 0.1853 after performing a Mann-Whitney U test. Concluding; it has been shown from this dataset that the used eyes-like stimulus did not have a significant effect on the average contribution levels, neither overall nor at the switching point from week 4 to 5.

Table 14



*H2.2: The average contributions will diminish during weeks 1 to 4 and during weeks 5 to 8 in Experiment 2*

As can be seen in Figure 7, the average contributions did decrease from week 1 to 4 and from weeks 5 to 8. Thus it can be concluded that it seems that the dual-process theory of response habituation has a stronger effect on the average contribution, than the neural activation in response to the detection of eyes-like stimuli, when these stimuli are being induced on the long run. However, it should be noted that eyes-like stimuli seemed to have a positive effect on the average contributions. In week 5, when the eyes-image was placed, it can be seen from the graph that the average contribution increased significantly. Specifically, the increase between average contributions of week 4 and 5 was from €0.42 to €0.59, constituting an increase of over 40%-points. However, this increase was still not sufficient to surpass the initial contribution level of €0.69 in week 1.

Figure 7

However, a Kruskal-Wallis test was performed in order to test if the medians of the 8 different samples came from the same population. This test indicates a significant p-value under the chi-squared distribution, if at least one median is significantly different from another one. This test was performed by pooling the daily data by weeks. From Table 15 it can be seen that this yielded a non-significant p-value of 0.4835 with a χ2 of 6.493 with 7 (k-1) degrees of freedom. Since this test uses ranks, it reported also the values with ties: p-value of 0.4700 and a χ2 of 6.615 with 7 degrees of freedom. Thus in this dataset the null hypothesis could not be rejected, indicating no significant differences between the 8 weeks. A possible reason for this statistical insignificance could be the fact that the standard deviations for each week were extremely high, as can be seen in Figure 8.

Table 15

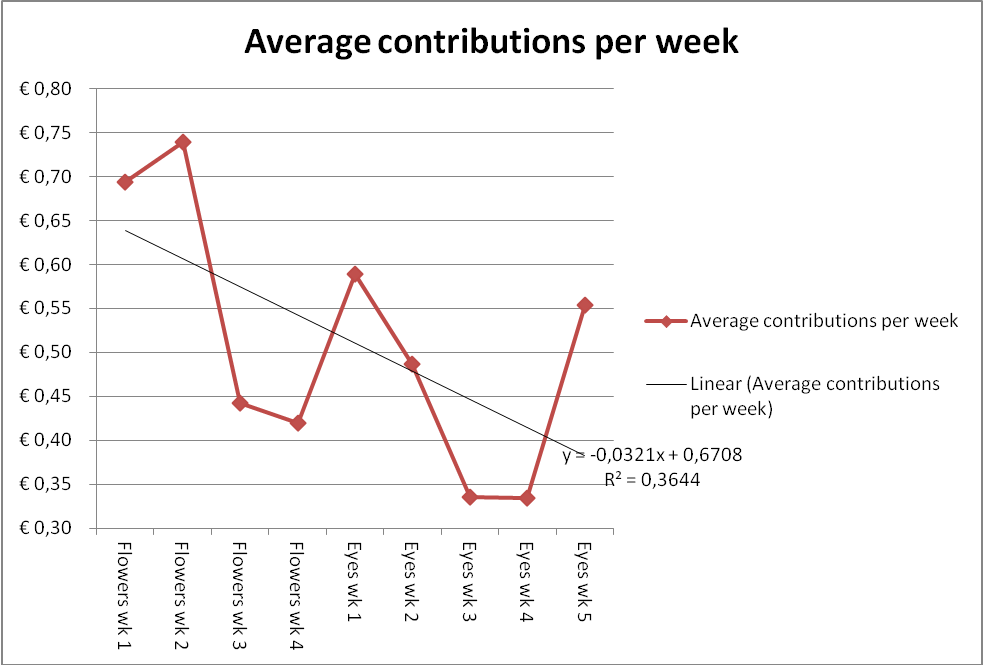


Figure 8

*H3.2: The average weekly contribution of week 9 will be higher than that of week 8 in Experiment 2*

As can be observed in Figure 6, the average contribution increased significantly from week 8 to week 9 (almost 67%, from €0.33 to €.55). However, the average contribution of week 9 still did not reach the initial average contribution levels of weeks 1 and 2, but almost reaching the effect eyes-image in week 5 (€0.59). However, it could be argued that alternation of eyes-images could have a positive effect on the average contribution levels. This is in line with the dual-process theory of habituation, where providing a stimulus first improves the dependent variable while decreasing this effect on the long run.

Figure 9



However, when statistically analyzing whether changing the eyes-like stimulus in week 9 did have an effect on the average level of contribution compared to week 8 using a Mann-Whitney U test, this effect was found to be non-significant with a z-score of -1.163 and a p-value of 0.2447, as can be seen in Table 16.

Table 16



For this experiment a power test was also performed ex post using an alpha of 5% and the current mean and standard deviation of both treatments. As can be seen in Table 17, this yielded an astonishing optimal sample of 1.156.828 for both treatments, due to the fact that the difference in the means of the two treatments was relatively small and only one price-level was used. By using a sample size of 55 subjects it is obvious that the power of test would be relatively low. From Table 18 it can be seen that this was 0.0529, thus only 5.29 percent of the null hypotheses were correctly rejected.

Table 17 Table 18

**

**Chapter 4: Discussion**

This paper set out to answer the questions whether the positive effect of eyes-like stimuli could be replicated in two field experiments in public goods games, and if so, how this effect develops through time on the average contribution levels. As was seen in the previous chapter; neither one of the experiments yielded significant results on the positive effect of eyes-like stimuli. In this chapter the attention will be focused on possible explanations and venues for future research in this area.

*Discussing Experiment 1;* although the setup of this experiment fairly resembled the experiment by Bateson et al. (2006), no significant positive effects of eyes-like stimuli could be detected. This gives reasons to believe that tag-based cooperation (Riolo et al., 2001) did not hold in this setting, even though the subjects all shared sufficient similar heritable tags as musicians. The limited cooperation in this experiment might be better explained by conditional cooperation where the strategy of limited contributions to the public goods was apparently enough to sustain the supply of the public goods.

This could have been due to the fact that the subject pool was different and/or the fact that no previous price-list had been used prior to this experiment. Regarding the subject pool, it might be the case that income-level effects attributed to decision-making-process whether to contribute (fully) or not. A priori it was argued that musicians received their income dispersed through the month instead of once a month. Moreover, musicians visiting the studios ranged from amateurs to professionals possibly implying different income-levels due to e.g. their musical skills, bargaining position as a result of which instrument they played, and reputation. Thus, future research might focus on the effect of eyes-like stimuli in settings where the income-level is further analyzed. Due to the nature of this experiment, this was impossible in this context.

Regarding the introduction of the price list, it could be argued that this implementation might have had a negative effect the average contributions. This due to the fact that prior to this experiment the contributions to the public goods were not explicitly asked from the subjects and depended only on their ‘voluntarism’. Introducing a price list might have crowded out this voluntary contributions and possibly induce cheating. However, no data was gathered before starting the experiment to being able to compare this with the average contributions after the price-list introduction. Recapitulating the previous positive findings in public goods games, as discussed in Chapter 1, it should also be noted that the eyes-like stimuli were introduced without altering the existing settings. For example; Bateson et al. (2006), where the use of a price-list in combination with the public goods was already in place, Ekström (2012), where the choice of either donating or cashing the money after returning bottles was already in place, and Powel et al. (2012), where the cans for donation was only adjusted by placing eyes-stickers on them. Thus, possible future research could experimentally analyze whether the introduction of an explicit pro-social behavior accompanied with eyes-like stimuli would yield significant results.

The nature of the common goods and the level of anonymity in this experiment are argued to be similar to that of Bateson et al. (2006), and thus left from this discussion. However, regarding the level of anonymity, it could be argued that due to the unobservability of the actual behavior of the subject, relevant information was lost. This caveat might be solved in future field experiments by introducing a hidden camera registering the actions and reactions of subjects on eyes-like stimuli. Please see Appendix E for a non-scientific example.

Despite the change in the setup in Experiment 1, this change in the setup raised the question on the positioning of these cues of being watched. As could have been seen in Figure 3, there was a relatively large increase in the average contributions between weeks 7 and 8, corresponding to the change in experimental setup. As can be seen in Picture 3 in the Appendix, shifting the image from just above eye-level to below eye-level just near the coffee-machine, might have had a positive effect on average contributions, since now the price-list including eye-image could have been relatively easier noticed.

Possible venues (for neuroscience) to explore might be possible costs of being watched as it seems that eyes-images might also induce feelings of discomfort.

*Discussing Experiment 2;* the design of this experiment also resembled that of Bateson et al. (2006) regarding the nature of the common goods and the level of anonymity. However, no significant effect was found of the eyes-like stimuli on average contribution levels (neither between the switch of images in week 5 nor between the overall average contributions in the two treatments). The observed cooperation might have been the result of a conditional cooperation based on learning. This reasoning is based on the findings in this dataset where it can be observed that the weekly average contribution levels decline in time. This could be the result of learning, as the subjects might have come to realize that free riding was not punished.

The fact that there was no significant effect of eyes-like stimuli could have been due to large standard deviations caused by the relatively small sample size (as was calculated with the power tests). Future research might focus on the analysis of static cues of being watched in a field experiment with a relatively larger sample size. However, other possible explanations might be at hand if the classifications described in Chapter 1 are analyzed in the light of this experiment relative to the other studies discussed earlier on cues of being watched in public goods games.

Regarding the subject pool; it can be argued that using a subject pool covering more than one cultural cluster as described by Inglehart & Baker (2000), might have affect the findings on the effect of eyes-like stimuli. Acknowledging the fact that in this paper these clusters were only marginally represented, it might have raised the awareness that interpreting the positive findings on eyes-like stimuli in previous public goods games might need some nuances. However marginally represented in this sample, it should be noted that especially the subjects not belonging to the ‘historically protestant’-cluster made use of the laundry facilities. This is due to the reason that Dutch students (belonging to the ‘historically protestant’-cluster) are more likely to have their parents living in the Netherlands and take their laundry with them upon visiting their parents in the weekends (whilst students and young professionals from abroad are more likely to use these facilities due to the lack of this alternative). Thus, future research might be able to focus on the use of a diversified cluster of subjects in experiments analyzing the effects of eyes-like stimuli in public goods games in order to assess to which extent these findings hold in different cultural clusters.

Another discussion point is the use of static eyes-like stimuli in public goods games. This experiment yielded no significant effects while adopting static eyes-like stimuli; however, as mentioned in Chapter 1, from the literature it seems that average contributions tend to decline over time for classical repeated public goods games. Additionally, the dual-process theory of response habituation (Groves & Thompson, 1970; Rankin et al., 2009) states that repeated application of a stimulus might first show an increase in responsiveness followed by a progressive decrease in some parameters of a response to an asymptotic level. Thus, a possible venue for future research might analyze the difference between a classical repeated public goods game and one with a static eyes-like stimulus (starting from the first independent observation) ceteris paribus. By doing so, e.g. the upper and lower bounds could be compared and analyzed in order to assess the degree of the effect of an eyes-like stimulus on the long run.

By using the data from the online washing-reservations schedule it was possible to calculate a proxy for the amount of subjects in the laundry facility on any given day divided by gender. Also the frequency of visits to this facility could also be approximated by the amount of washing reservations on any given day, since one subject could make more than one reservation a day. In order to assess the possible multiplier effects as stated by Ekström (2012), the amount of overlapping reservations between different subjects served as a proxy for interaction on any given day. Since no lower contributions than zero could and have been made (no stealing), the daily average contributions were censored at this lower limit of zero. For the regression of these independent variables on the dependent variable (daily average contributions), a Tobit models was used. A dummy variable for treatment level was incorporated, which took the value of zero for treatment 1 and the value of 1 for treatment 2. This regression yielded only a significant negative coefficient for ‘males’ of -0.3926128 with a p-value of 0.008 and a significant treatment-effect of 0.6372737 with a p-value of 0.042 (significant at 5%). This model was omitted from the results since; no information was available whether this gender effect was caused by one, a couple or all the male subjects. Moreover, conducting this censored regression on only treatment level yielded no significant effects. However, this raised the possibility for future research to analyze whether gender effects might play a role in the distribution of the contributions in a repeated public goods game.

**Literature**

* Ariely D., The honest truth about dishonesty, Harper Collins Publishers, 2012
* Andreoni J., 1988. Why free ride?: Strategies and learning in public goods experiments. *Journal of Public Economics*, volume 37, issue 3,, pages 291-304.
* Andreoni, J., & Petrie R., 2004. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, volume 88, issue 7-8, pages 1605-1623.
* Arnett, J., 2008. The neglected 95%: Why American psychology needs to become less American. *Am. Psychol.* 63**,** pages 602–614.
* Abbink, K., & Herrmann, B., 2010. The moral costs of nastiness. *Economic Inquiry*, 49, pages 631-633.
* Baillon, A., Selim, A., & van Dolder, D., 2013. On the social nature of eyes: The effect of social cues in interaction and individual choice tasks. *Evolution and Human Behavior*, 34, pages 146-154.
* Bar-Hillel, M., 1973. On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 9, pages 396-409.
* Bateson M., Nettle, D., & Roberts, G., 2006. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, volume 2, no.3, pages 412-414.
* Bohm, P., 1972. Estimating demand for public goods: An experiment. *European Economic Review*, volume 3, issue 2, pages 111-130.
* Burnham, T.C., & Hare, B., 2007. Engineering human cooperation. Does involuntary neural activation increase public goods contributions? *Human Nature*, volume 18, pages 88-108.
* Camerer, C.F., 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
* Cookson, R., 2000. Framing Effects in Public Goods Experiments. *Experimental Economics*, volume 3, pages 55-79.
* Davis, D.D., & Holt, C.A., 1993. *Experimental economics*. Princeton, NJ: Princeton University Press.
* Ekström, M., 2012. Do watching eyes affect charitable giving? Evidence from a field experiment. *Experimental Economics*, volume 15, pages 530-546.
* Ellsberg, D., 1961. Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, volume 75, pages 643-669.
* Emery, N.J., 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, volume 24, issue 6, pages 581-604.
* Ernest-Jones, M., Nettle, D., & Bateson, M., 2011. Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior*, volume 32, pages 172-178.
* Fehr, E., & Fischbacher, U., 2004. Social norms and human cooperation. *Trends in Cognitive Sciences*, volume 8, issue 4, pages 185-190.
* Fehr, E., & Schneider, F., 2010. Eyes are on us, but nobody cares: are eye cues relevant for strong reciprocity? *Proceedings of The Royal Society B: Biological Sciences*, volume 277, pages 1315-1323.Fischbacher, U., Gächter, S., & Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Economic Letters*, volume 71, issue 3, pages 397-404.
* Gächter, S., & Fehr, E., 1999. Collective action as a social exchange. *Journal of Economic Behavior and Organization*, volume 39, pages 341-369.
* Gächter, S., & Herrmann, B., 2009. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of The Royal Society B: Biological Sciences*, volume 364, no. 1518, pages 791-806.
* Gächter, S., Herrmann, B., & Thöni, C., 2010. Culture and cooperation. *Philosophical Transactions of The Royal Society B: Biological Sciences*, volume 365, pages 2651-2661.
* Gintis, H., Bowles, S., Boyd, R., & Fehr, E., 2003. Explaining altruistic behavior in humans. *Evolution and Human Behavior*, volume 24, issue 3, pages 153-172.
* Groves, P.M., Thompson, R.F., 1970. Habituation: A dual-process theory. *Psychological Review*, volume 77, pages 419-450.
* Haley, K.J., & Fessler, D.M.T., 2005. Nobody’s watching? Subtle cues can affect generosity in an anonymous economic game. *Evolution of Human Behavior*, volume 26, pages 245-256.
* Hare, B., Call, J., & Tomasello, M., 2001. Do chimpanzees know what conspecifics know? *Animal Behaviour*, volume 61, issue 1, pages 139-151.
* Hare, B., Call, J., & Tomasello, M., 2006. Chimpanzees deceive a human by hiding. *Cognition*, volume 101, pages 495-514.
* Haxby, J.V., Hoffman, E.A., & Gobbini, I., 2000. The distributed human neural system for face perception. *Trends in Cognitive Science*, volume 4, pages 223-233.
* Henrich, J.H., Heine, S.J., & Norenzayan, A., 2010. Most people are not WEIRD. *Nature*, volume 466, page 29.
* Herrmann, B., Thön, C., & Gächter, S., 2008*a*. Antisocial Punishment Across Societies. *Science*, volume 319, no.5868, pages 1362-1367.
* Inglehart, R., & Baker, W., 2000. Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, volume 65, no.1, pages 19-51.
* Isaac, R.M., Walker, J.M., & Thomas, S.H., 1984. Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, volume 43, pages 113-149.
* Johnson, N.D., & Mislin, A.A., 2011. Trust games: A meta-analysis. *Journal of Economic Psychology*, volume 32, issue 5, pages 865-889.
* Keser & van Winden (2000), Keser, C., & van Winden, F., 2000. Conditional Cooperation and Voluntary Contributions to Public Goods. *The Scandinavian Journal of Economics*, volume 102, issue 1, pages 23-29.
* Ledyard, J., 1995. Public Goods Experiments: A Review. In A. Roth and J.Kagel (eds.), *Handbook of Experimental Economics*. Princeton University Press.
* Masclet, D., Noussair, C., Tucker, S., & Villeval, M.C., 2003. Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *The American Economic Review*, volume 93, no.1, pages 366-380.
* Milinski, M., Semmann, D., & Krambeck, H., 2002. *Proceedings of The Royal Society B: Biological Sciences*, volume 269, no.1494, pages 881-883.
* Nordstokke, D.W., & Zumbo, B.D., 2010. A New Nonparametric Levene Test for Equal Variances. *Psicológica*, volume 31, pages 401-430.
* Oda, R., Niwa, Y., Honma, A., & Hiraishi, K., 2011. An eye-like painting enhances the expectation of good reputation. *Evolution and Human Behavior*, volume 32, pages 166-171.
* Powell, K.L., Roberts, G., & Nettle, D., 2012. Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket. *Ethology*, volume 118, pages 1011-1096.
* Raihany, N.J., & Bshary, R., 2012. A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B: Biological Sciences*, volume 279, pages 3556-3564.
* Rankin, C.H., Abrams, T., Barry, R.J., Bhatnagard, S., Clayton, D.F., Colombo, J., Coppola, G., Geyer, M.A., Glanzman, D.L., Marsland, S., McSweeney, F.K., Wilson, D.A., Wu, C.F., & Thompson, R.F., 2009. Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory*, volume 92, issue 2, pages 135-138.
* Rege and Telle (2004) Rege, M., & Telle, K., 2004. The impact of social approval and framing on co-operation in public good situations. *Journal of Public Economics*, volume 88, pages 1625-1644.
* Rigdon et al. (2009) Rigdon, M., Ishii, K., Watabe, M., & Kityama, S., 2009. Minimal social cues in the dictator game. *Journal of Economic Psychology*, volume 30, pages 358-367.
* Riolo, R.L., Cohen, M.D., & Axelrod, R., 2001. Evolution of cooperation without reciprocity. *Letters to Nature*, volume 414, pages 441-443.
* Soetevent, A., 2005. Anonymity in giving in a neutral context-a field experiment in 30 churches. *Journal of Public Economics*, volume 89, issue 11-12, pages 2301-2323.
* Sylwester, K., & Roberts, G., 2010. Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters*, volume 6, no.5, pages 659-662.
* Van Dijk, F., Sonnemans, J., & van Winden, F., 2002. Social ties in a public good experiment. *Journal of Public Economics*, volume 85, issue 2, pages 275-299.
* Wilkinson, N., 2008. An Introduction to Behavioral Economics. *Palgrave MacMillan*, pages 272-274.

**Appendixes**

***Appendix A***

*Classification of economic games*

When summarizing the above-mentioned positive findings of eyes-like stimuli on behavior, several aspects should be kept in mind. As Wilkinson (2008) stated, it is useful to classify the different types of games according to certain important characteristics. I will briefly summarize these characteristics with respect to the above-mentioned findings:

1. Cooperative and non-cooperative games: In cooperative games the subjects can communicate with each other and thus base their decisions upon agreements. However, much of this type of activity is explicitly prohibited by law in developed countries. Thus most of the economic games are of the non-cooperative kind where subject involve in forming self-enforcing reliance relationships.
2. Two-player and multi-player games: Having more players tends to increase the likelihood of defection, particularly in the ‘one-shot-games’. With more players it is important to defect before others do and this can be prevented only if the defectors can be easily detected and punished by other subjects.
3. Zero-sum and nonzero-sum games: Zero - sum game is when the gain of one player(s) is by definition the loss of another player(s). However, even when monetary gain and losses offset each other (nonzero-sum games); the utilities of such gains and losses might not offset each other due to loss-aversion (Kahneman & Tversky, 1979).
4. Perfect and imperfect information: When players know for certain all the payoffs for each strategy, including their own and that of other subjects, the game is classified as having perfect information.
5. Static and dynamic games: Static games involve simultaneous moves where the move of the other subject is not known. Dynamic games involve sequential moves where the move of the first subject is known to the second subject prior to making his/her decision.
6. Discrete and continuous strategies: This reflects on the size of the array of possibilities that are presented in each game. Subjects may have to choose between either two or more options.
7. ‘One-shot’ and repetitive games: Where repetitive games entail the continuous interaction among subjects thus giving them the opportunity to change their decision (variables) at regular intervals. In a ‘one-shot’ game, where the time interval is relatively large(r), the decision making process may be repeated in the future. However, due to a possible changed scenario this might involve quite different (individual) payoffs.

Applying this classification, a public goods game can be described[[41]](#footnote-41) as a multi-player -, non-zero sum -, static -, and repetitive game with discrete strategies in a (non-)cooperative setting with imperfect/perfect information[[42]](#footnote-42). A dictator game, from the proposers’ view, can be described as a non-cooperative-, two-player -, zero-sum -, and ‘one-shot’-game with discrete strategies. Regarding the information, this depends on the level of anonymity in each game. This game is neither dynamic nor static in this context, since the reaction of the responder has no direct influence on the decision of the proposer, except possibly indirect influences due to altruism, and reputational/punishment concerns. From the responders view the dictator game can be described the same as from the proposers’ view, except for the lack of cooperation levels and discrete/continuous strategies, since the responder can only accept the decision of the proposer. As mentioned before the effects of eyes-like stimuli are bound to psychological effects and that even without social cues, subjects tend to violate the predictions of standard economic models (SEM).

**Appendix B**

*Communication to the subjects of Experiment 2*

Dear Residents,

As you may know, the vending machine has been causing some people to pay without receiving their product(s). In order to prevent this, the vending machine will be out of order starting Monday 29th of April 2013 till further notice. Since this has proven to be a desired service, an alternative will be offered to you starting on **Monday 29th of April** till further notice.

In the laundry room on the first floor you will find a basket with 2 cans of Coca-Cola, one can of Fanta, one bag of Croky Chips Naturel, one bag of Croky chips Bolognese, one bar of Snikcers and one bar of Bounty. This basket will be refilled daily to the above mentioned proportions. Besides this basket you will also find an ‘honesty box’ in which you can deposit your money after having taken one or more products. The prices of all these products are set at €0.80 per item.

Kind regards,

Royston

**Appendix C**

*Communication to the Housing Committee regarding Experiment 2*

Dear Colleagues,

Regarding the problems we have been facing with the vending machine I propose a solution. I propose to disconnect the vending machine and offer the Residents a basket with similar products which will be place in the laundry room. Prices will be held at the same level as the current ones in the vending machine. Since there is no control, I will rely on the reciprocity of the Residents to pay for the product(s) taken from the basket. Along with this basket an ‘honesty box’ will be placed where they can deposit the money. This will of course be locked and will be emptied daily by me (after refilling the basket). Any losses incurred due to these actions will be for my personal account.

This solution is two-folded since, apart from providing a better service, for once we might be able to gather real information regarding the revenues incurred by the vending machine company.

Starting from Monday 29th of April I will adopt this method for a couple of weeks. Needless to say, I will keep you informed regarding the revenues at our monthly meetings.

Kind regards,

Royston

**Appendix D**

*Communication to the subjects regarding ending of Experiment 2*

Dear Residents,

As of today, Monday 1st of July, the vending machine will be working again.

The company has gone to the lengths of fixing the machine so you can continue enjoying your drinks and snacks.

Kind regards,

Royston

**Appendix E**

*Pub trivia: application of a public goods game in a marketing campaign*

Concluding my thesis I would like to leave you with the social “experiment” conducted by HONEST Tea (2013); regarding honesty of Americans on 61 locations throughout all 50 states from July 8- 18, 2013[[43]](#footnote-43). The company setup unmanned kiosk stocked with its beverages and offered them for $1/bottle on the honor system. This honor system was a transparent glass cube on a stand with just enough space to deposit the dollar bill. Subjects would not suffer any consequences from the company if they took their products without paying, since the company did its best to induce the feeling that no one was watching. However, due to the use of public locations (after all it was a marketing campaign), subjects were not entirely anonymous to the gazes of others or strangers. Moreover, employees were (unobtrusively) stationed near the kiosks in order to gather the data. Thus, subjects were on no occasion alone and fully anonymous in their decision, since at some point (however brief this might have been), the employees would have had to look in the direction of the subject in order to gather the data on contribution behavior. However, they did their best to be as consistent as possible and treated every location and data gathering the same way. Moreover, as the company states, this a light-hearted social experiment rather than a controlled scientific research test[[44]](#footnote-44).

They found an overall average percentage of honesty across all states of 92%, with a range between states of 80-100%. In previous smaller-scale campaigns they found an honesty range of 86-99% for American cities in 2011, respectively 79-100% in 2012, with an overall honesty average of 93% in 2012.

Figure 10 (Data Experiment 1)



Figure 11 (Data Experiment 2)

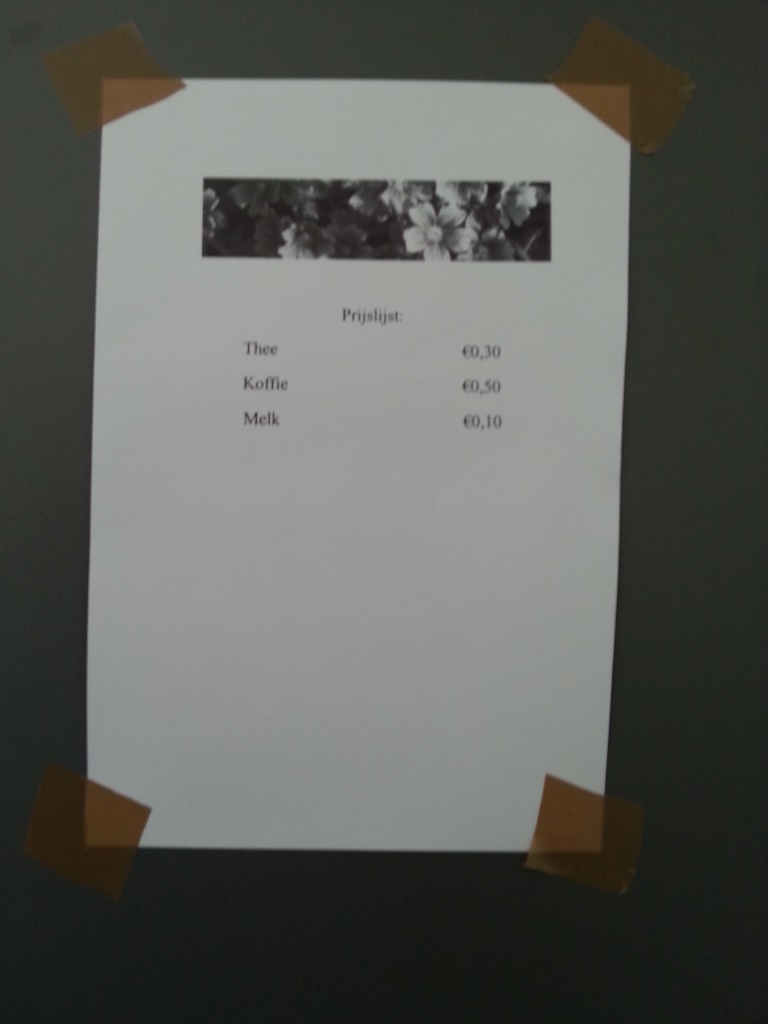


Figure 12 (Cost overview Experiment 2)

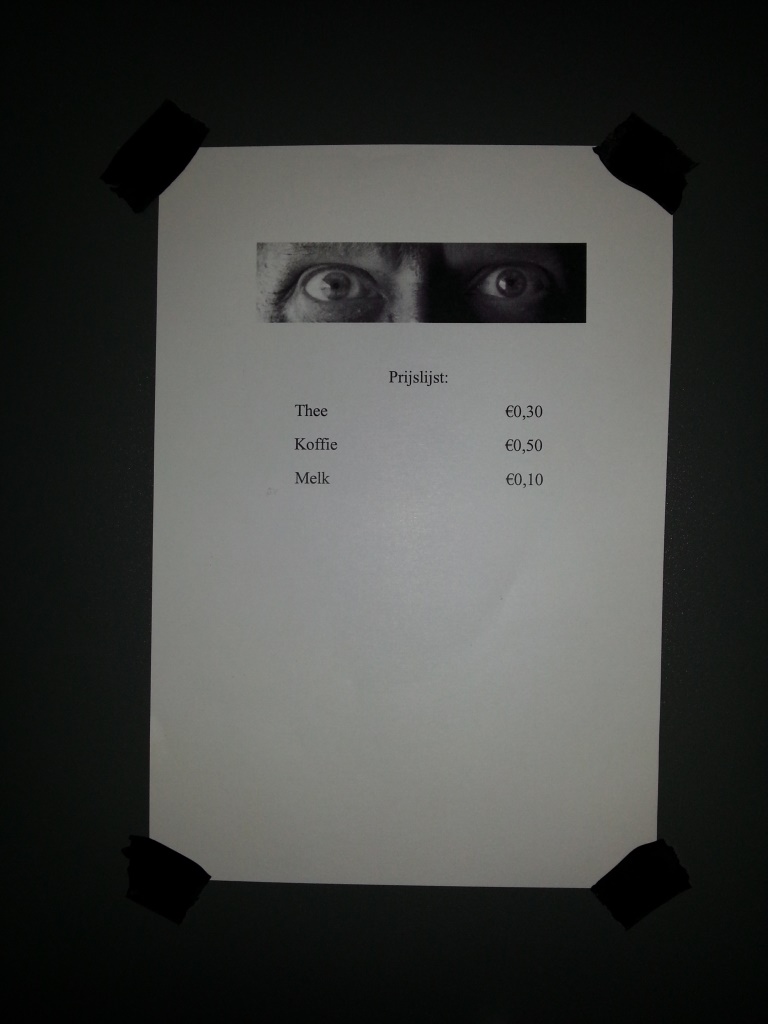


Picture 1 Setup Experiment 1 (Treatment 1)





Picture 2 Setup Experiment 1 (Treatment 2)



Picture 3 Setup Experiment 1 (Change setup)



Picture 4 Setup Experiment 1 (Treatment 3)



Picture 5 Setup Experiment 2 (Treatment 1)





**Picture 6 Setup Experiment 2 (Treatment 2)**



Picture 7 Setup Experiment 2 (Treatment 3)



1. This eyes-image originates from the study by Bateson et al. (2006), whom approval have been given by e-mail. [↑](#footnote-ref-1)
2. Please see Ariely (2012) for experiments conducting on cheating behavior illustrating that cheating increased when the pay-offs where in tokens rather than money, even when the table for changing the tokens into hard cash was a couple of steps away. However, in this part I will stick to the classical experiments using money. [↑](#footnote-ref-2)
3. Classic: in the sense that no eyes-like stimuli were used in these researches. [↑](#footnote-ref-3)
4. These views are based on the papers by: Isaac et al. (1984), Andreoni (1988), Keser & van Winden (2000), Fischbacher et al. (2001), Riolo et al. (2001), Milinski et al. (2002), and Sylwester & Roberts (2010), along with some personal additions. Please refer to the papers for their original findings. [↑](#footnote-ref-4)
5. Please see Appendix A for the reasoning behind the classification of these games. [↑](#footnote-ref-5)
6. In this context: A type of empirical economic study where real or perceived decisions of subjects are being measured in a controlled setting where they know they are participating in one. [↑](#footnote-ref-6)
7. The 32 days consist of; 14 days in the congruent condition (7 eyes and 7 flowers) and 18 days in the incongruent condition (9 eyes and 9 flowers). [↑](#footnote-ref-7)
8. For every additional party member the odds ratio was 1.48 [↑](#footnote-ref-8)
9. For every additional person in the cafeteria the odds ratio was 0.99 [↑](#footnote-ref-9)
10. Obviously more factors can be distinguished in order to classify these findings; however, for the scope of this study this has been limited to four. The framework by Wilkinson 2008 was used for this classification; please see Appendix A for a brief outline of this framework. [↑](#footnote-ref-10)
11. The division was 3 boxes with stickers of eyes and 3 boxes with stickers of three stars. Furthermore, these boxes were alternately placed on checkouts and thus, eyes and control images were evenly distributed across the width of the store. [↑](#footnote-ref-11)
12. That is, if the increase in the last week is disregarded. [↑](#footnote-ref-12)
13. I acknowledge the fact that this survey was not done on economic journals, but a quick review of the subject pools used in economic experiments yields the same notion: mostly economic undergraduate students are being used. Thus, these percentages are merely for grasping the scope of this phenomenon. [↑](#footnote-ref-13)
14. However, they did not find a significant effect for education level and income level on donations. A remarkable finding was that they found a significant positive effect of images of flowers on donations rather than images of eyes. [↑](#footnote-ref-14)
15. An economic experiment where subjects are endowed with a certain amount of cash (the proposers), after which they have to decide upon sharing with another (unknown) subject (the responder). Several variations are possible concerning e.g. the endowment, anonymity and the settings. [↑](#footnote-ref-15)
16. [↑](#footnote-ref-16)
17. Positioned in the form of a ‘V’ [↑](#footnote-ref-17)
18. In principle the same as a dictator game, however, the responder now has the choice to either accept or reject the offer made by the proposer. In case of rejection, both players receive a zero payoff. [↑](#footnote-ref-18)
19. I am grateful to Drs. Van Dolder for conceptualizing my thoughts by introducing me to this terminology. [↑](#footnote-ref-19)
20. Keilestraat 3, 3029BA, Rotterdam, the Netherlands. [↑](#footnote-ref-20)
21. <http://www.duo.nl/particulieren/klantenservice/betaaldata.asp> [↑](#footnote-ref-21)
22. <http://www.arbeidsrechter.nl/wanneer-loon-betalen-overmaken-welk-tijdstip> [↑](#footnote-ref-22)
23. Please note that I am referring to automated machines in which you have to deposit coins in order to get the desired product and not settings with an honesty box in place. As reference were used automats in public settings. However, the prices were lowered in accordance with the building manager, since subjects used porcelain cups which had to be washed and dried after use along with the used utensils. [↑](#footnote-ref-23)
24. I personally emptied the box every week after measuring the amounts consumed, and hereafter I placed everything back exactly as I found them. Good care was taken that no one could see me during these actions, in fact no one did. [↑](#footnote-ref-24)
25. Grams instead of liters, since in this setting powder-milk is used instead of liquid milk. [↑](#footnote-ref-25)
26. This will be done by deducting the total amount of the previous week with the total weighted amount in the current week. For the weighting I will be using a kitchen appliance for measuring of the brand ‘Soehnle’, type ‘Page’, art.-no; 66108, which has a range of 0-9999 grams displayed in full numbers (no fractions). [↑](#footnote-ref-26)
27. The average cost price of these products was €0.65, calculated by dividing the total costs of this experiment by the amounts of goods taken. The total costs consist of the purchase of the; honesty box, beverages, snacks, chocolates, basket, and gas to and from the wholesaler. Please see the Figure 7 in the Appendix for an overview of the costs, revenue and ‘profit’ made. All goods were purchased at a wholesaler. [↑](#footnote-ref-27)
28. Sometimes it issued the products, sometimes not, sometimes you got your change back, sometimes not. [↑](#footnote-ref-28)
29. Please see Appendix D for the communication note to the subjects regarding this change. [↑](#footnote-ref-29)
30. Since it is customary to conduct the communication in English, no translation was needed. Please see Appendix B for this note. [↑](#footnote-ref-30)
31. This group contained 51 of the 56 residents, however, since I am also a resident in the building the actual ratio would be 50/55. [↑](#footnote-ref-31)
32. Each resident automatically received an e-mail after the posting on this website; moreover, downloading information from the website was also restricted to only residents. [↑](#footnote-ref-32)
33. Every resident/subject is prone to the use of this entrance at least once a week since this is also where their mailbox is situated. Moreover, using this entrance facilitates the use of the elevator which they use to transport their bikes to the underground storage room. [↑](#footnote-ref-33)
34. Unfortunately this could not be controlled for, since every resident by default possesses two keys which can open all the locks of the doors to the general areas/facilities besides their apartment door. However, to my knowledge, most residents opt to store their second/spare key with a friend or relatives. Thus, it could be argued that these spare-key-holders have more or less the same traits as the residents. Moreover, it need not be necessary that as a result of storing the spare-key that they visit the building frequently. [↑](#footnote-ref-34)
35. Please see Figure 5 in the Appendix for an illustration. Upon the daily emptying of the honesty box I checked whether this tape was still effective and that at least one coin still could get stuck upon turning the box. It turned out that no one ever tried to steal any coins, at least no evidence was found throughout the experiment. [↑](#footnote-ref-35)
36. This is concluded from the fact that residents have to buy washing and drying coins which I have to deliver to them. [↑](#footnote-ref-36)
37. Namely my method of managing; they have come to realize that I like to take up new challenges. Resulting in for example; being the first building of the building corporation to implement a fully digitalized washing-reservation-schedule, and the online purchase of washing and drying coins. This was done despite the fact that the building corporation gave a negative advice (non-binding) on this project. This project turned out to be a success and up until today we are still helping other buildings with their digital-projects after reference of the building corporation. (Please do not interpret this as bragging but a mere illustration of my point, well ok, maybe some overconfidence). [↑](#footnote-ref-37)
38. This is truly a coincidence, since this rule was already in the making before I even started with this research. [↑](#footnote-ref-38)
39. This refers to gathered knowledge over the years as a building manager. [↑](#footnote-ref-39)
40. Moreover, even after adding the data points from week 9 this was insignificant with a p-value of 0.1141 and a z-score of 2.93. [↑](#footnote-ref-40)
41. In the context of the above-mentioned games, described in the literature review. [↑](#footnote-ref-41)
42. This is dependent on the level of anonymity. [↑](#footnote-ref-42)
43. <http://www.honesttea.com/news/pressreleases/> [↑](#footnote-ref-43)
44. <http://thenationalhonestyindex.com/default/index/about> [↑](#footnote-ref-44)