GROUP COMPOSITIONS IN THE PUBLIC GOODS GAME

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis: Behavioural Economics – Market & Organisation track

Name student: Reza Safarzade 412695

Supervisor: Xiao Yu

Second assessor: P.P. Wakker

Date final version: 18-12-2020

Abstract

This study investigates the effect of group composition in combination with the default in the public goods game with the following research question: *"What the effect is of group composition on the success rate of the default in the public goods game?"*. Participants are put randomly in one out of four public goods games. Two out of four are in a diverse group, both genders and multiple nationalities, and the other two are homogenous. I.e., all-white and all-men. The default option is present in one of each different group composition. Results show that the average contribution in the homogenous group is €6.22 and in the diverse group €7.21. This difference is significant, with a P-value of 0.0628. Receiving the default in the diverse group compared to receiving it in the homogenous group increases contribution from €6.08 to €7.76 with a P-value of 0.0144. This difference comes mostly from men who significantly increase their contribution from €5.53 to €8.07 with a P-value of 0.0223. Regression analysis confirms this. Being in a diverse group significantly increase the contribution by 1.369. Being in the default – diverse group significantly increase the contribution by 1.717. This experiment shows that the social environment and the default are complementary to each other. This study emphasizes the importance that changes in choice architecture cannot be without studying the target audience and the social environment.

Table of Contents

Introduction

There is increasing awareness of the importance of sustainable choices. Current consumption is not sustainable due to the limited amount of resources and pollution. It is becoming common knowledge that current consumption is not sustainable. However, still too many people do not choose sustainable. It is the responsibility of the policymaker to educate and guide people towards choices that are better for the environment. Multiple actions are possible to change this. Lehner, Mont & Heiskanen (2016) mention that currently, the most widely used policy is to provide information to promote sustainable consumption. However, this assumes the rationality of humans. Tversky & Kahneman (1974) show that it is not possible to trust this rationality completely. People have bounded rationality. Choices are biased and reached with heuristics. This bounded rationality makes it essential for the policymaker to step in. People still need the freedom to choose otherwise. Thaler & Sustein (2003) call this libertarian paternalism. "An approach that preserves freedom of choice but that authorizes both private and public institutions to steer people in directions that will promote their welfare" (Thaler & Sustein, 2003).

Previous studies have shown the positive effect of nudges. Nudge tools are seen as a complement to the traditional policy instruments rather than as a substitute for laws and regulations and economic tools (Lehner, Mont & Heiskanen, 2016). For example, in order to fight food waste, some restaurants and hotels provided their customers with smaller plates and successfully reduced food waste (Kallbekken and Sælen, 2013) (Freedman and Brochado, 2010). Minor adjustments in the choice architecture can have a significant impact. Goldberg and Gunasti (2007) show, for example, that placing fruits close to the register instead of sweets will cause people to buy more fruit.

However, the choice architect needs to understand the target audience and the social environment of choices. While the intention might be right, it can be counter-effective. Lombardini & Lanoski (2013) show this with the following example. A school in Finland forced a vegetarian day for all pupils. The goal was that meat consumption would decrease and pupils would eat vegetarian for at least one day. However, what happened is that pupils left school property to get lunch elsewhere. Pupils created their freedom of alternative elsewhere. Besides, food waste increase because pupils would not finish their meals. Forcing people will not give the desired result. There needs to be the indifference of choice or leaning preference towards the nudge. Scaling up nudge interventions will likely prove disappointing unless sound knowledge of the target audience and the behavioural environment are available. (Lehner, Mont & Heiskanen, 2016).

The nudge of interest in this study is the default. Default is the condition that is imposed when an individual fails to make a decision (Johnson & Goldstein, 2003*).* I.e. without actively taking action, the choice is made. Nevertheless, there is still the freedom to choose otherwise. Studies show the default effect in environmental choices. Egebark & Ekström (2016) change the default in printing from one-sided to double-sided. This minor adjustment shows a 15% decrease in paper consumption. Default rules can have an exceedingly large impact on environmental quality. This could potentially be larger than that of significant economic incentives or serious efforts toward moral suasion or environmental education (Sunstein & Reisch, 2014).

Thinking requires cognitive effort, and that is costly. Defaults are therefore also seen as a recommendation by the choice architect. Sticking with the default seems to allow one to bypass a stressful and awkward decision (Pichert & Katsikopoulos, 2008). It is not unreasonable to stick with the default because the default could be interpreted as a recommendation from policymakers and indicates socially desired behaviour (Johnson & Goldstein, 2003; McKenzie, Liersch & Finkelstein 2006). (Pichert & Katsikopoulos (2008) show one of the most promising natural experiments. A small German town took the opportunity to control their electric grid. The small majority won. 52% versus 48% that favoured taking control and putting an environmentally friendly provider in place. People still had the freedom to choose otherwise. The green option became the default. Almost all electric meters are connected to the green option after eight years. One of the primary advantages of green defaults is that they can have beneficial effects while maintaining freedom of choice and hence respect for heterogeneity (Sunstein & Reisch, 2014).

Recommendations are also useful when dealing with complex decisions that require a vast amount of knowledge and expertise. Many people interpret the default option as an implicit recommendation in the sense that if there is a standard established by authority, it cannot be (very) wrong (Pichert & Katsikopoulos, 2008). An essential factor for the success of the default is trust. Evans, Dillon, Goldin & Krueger (2011) emphasize in their work the importance of trust required in setting the default option. To understand the default, understanding of trust is essential.

Lehner, Mont, & Heiskanen (2016) emphasize that it is essential to understand the target audience and the behavioural environment. Without this knowledge, the nudge will not be sufficient. Target audience means not only the decision-maker but also the social environment surrounding the decision-maker, e.g. group composition. Various studies show the gender effect in economic experiments. While Eckel & Grossman (1998), Andreoni & Vesterlunder (2001), Nowel & Tinkler (1994) find that women

behave more altruistically compared to men, Brown-Kruse & Hummels (1993) find the opposite. Namely, men behave more altruistically. There is yet to find a clear answer for gender effects in economic experiments.

Nevertheless, there is no clear answer to the effect of group composition on contribution and adherence to nudges in the public goods game. This research contributes to the current literature new insights on behaviour in different social environments and the adherences to nudges in these different group compositions. The contribution consists of the answer to the following research question: *"What the effect is of group composition on the success rate of the default in the public goods game?"*. Literature formulates the following sub hypotheses. Confirming or falsifying these helps to answer the research question.

H1: *"Contribution will be higher in the default group compared to the non-default group."*

H2: *"Contribution in the diverse group is higher than in the homogenous group."*

H3: "Contribution *is higher with the default option compared to no default within their respective group composition."*

This experiment has four public goods games to answer the research question. In two, there is a diverse group composition. I.e., multiple genders and ethnicities. The other two have a homogenous group composition. I.e., all-male and all-white. In each of these group compositions, there is a default present with contributing the maximum amount. Participants randomly play one out of four games. Nonparametric tests compare samples. Regression analysis provides robustness.

Results show some significant and insignificant results. Average contribution in the diverse group compared to the homogenous group is significantly higher, €7.21 vs €6.22 with P=0.0628. Similar results are when comparing diverse – default vs homogenous – default, €7.76 vs €6.08 with P=0.0144. In this same group, there is a significantly higher proportion of people sticking with the default. 0.50 vs. 0.297 with P=0.083. The complementary effect shows when giving the treatment of default in the diverse group—contribution increases from €6.48 to €7.76 with P=0.1044. The proportion of sticking with the default increases from 0.308 to 0.50 with P=0.1376. Additional robustness checks from regressions show a significant increase of €1.369 in contribution when playing in the diverse group and a significant increase of €1.717 in contribution when playing in the diverse – default group. Contribution decreases insignificantly when presenting the default in the homogenous group, €6.37 to €6.08 P=0.7462. The results show that a diverse group composition significantly increases contribution. Above that, the default effect

is present only in diverse groups. I.e. diverse group composition increases the success rate of the default. There is an interaction effect.

First, current literature about the topics is discussed. The literature research first shows current theories about the default and why there is the default effect, after the default current literature about gender effect and social diversity. After the literature research, this study shows the methodology and the results of the experiment. Results contain descriptive statistics, hypotheses testing, and regressions analysis. Afterwards, results are associated with the research questions, hypotheses, and current literature. The discussion contains implications, validity, and future suggestions.

Literature review

Default option

There is an increasing amount of policymakers using the default option. One of the most used examples is donor registration. Countries are changing their policy from being an opt-in to opt-out policy. I.e. the default changes from not being a donor to being a donor. Small adjustments impact choice architecture. Huh, Vosgerau & Morewedge (2014) mention three variants of the default. The first one mentioned previously is the opt-in policy. The second variant is a base model being the default, see Park, Jun, & MacInnis (2000). In this base model addition or subtractions are made. The third variant describes McKenzie, Liersch, and Finkelstein (2006). This variant does not show alternatives, and it is required to request and receive these. These variants have the same purpose. They are making the default the first consideration. Hence, becoming the status quo. “if there is a default, do nothing about it”. (Johnson & Goldstein, 2003). Why are these minor adjustments significant?

Multiple papers come up with their reasoning for the default effect (McKenzie, Liersch & Finkelstein, 2006; Johnson & Goldstein, 2003; Huh, Vosgerau & Morewedge 2014). Johnson & Goldstein (2003) narrow it down to three reasons. Deviating from the default incur costs and require cognitive effort. Decision-makers, and people in general, like to be cost-effective and follow the most efficient choice. The second reason is loss aversion. The default becomes the reference point, and deviating from this can incur losses. These losses are weighted heavier than potential gains, according to Kahneman & Tversky (1979). Lastly, the default acts as a recommendation. In the public goods game, the default acts as a reference point and recommendation. The following part only focusses on these two.

Reference point

The characteristic of the default is that the policymaker selects the option for the decision-maker. By being already selected, it becomes the reference point. Acting like the decision-maker already owns the default and alternatives are deviations is called instant endowment by Dinner, Johnson, Goldstein & Liu (2011). Observing alternatives as potential gains or losses creates preferences. Loss aversion makes losses weighted heavier. Besides, previous experience and memories assist the evaluation of alternatives. McKenzie & Nelson (2003) show the effect of memory on assessment. When the cup is empty first, the participant describes the cup as half-filled seeing a half-cup. The opposite is also true. When a cup is full first, it will be described as half empty when shown afterwards. This experiment shows that participants use retrievable information from memory describing present situations. This theory is built further with query theory by Dinner et al. (2011).

Query theory suggests that when confronted with the possibility to sell an object, people generate a list of why they will not sell the object. Afterwards, generating a list of reasons why they want to dismiss the object. Johnson, Haubl & Keinan (2007) confirm this. Their study shows that seller first list reasons to keep the object. This list of keeping is also more prominent than the list of selling. Buyers show the opposite behaviour. Reasons to keep the money is more significant than the list of reasons to spend.

Dinner, Johnson, Goldstein & Liu (2011) link this in turn with output interference. The process of retrieving and outputting information is output interference. I.e. when presented with an option, the first list generated consist of reasons why it is good. See Anderson, Bjork & Bjork (1994) for more. Linking these two theories together means that when presented with the default, people generate a list of reasons why the default is good. Dinner et al. (2011) test these generated lists with the following experiment.

Participants have to choose between two lightbulbs. Two groups form with each their default. One default is a greener choice than the other. Participants have to list their thoughts about the lightbulb. Afterwards, thoughts are either for or against the choice. Results show that there are a default effect and decision-makers first list-advantages, and this is in higher quantity compared to the alternative. This experiment shows that the default becomes the reference point. Participants act like they already own the default. Running the same experiment, but now participants list the reason after deciding brings similar significant results. The following experiment is a 2x2 design to find a causal effect.

The set-up of the second experiment is the same as the first. However, now there is a consistent and inconsistent situation. In the consistent situation first, the positive aspects of the default are listed and negative of the alternative. Afterwards, they list the negative aspects of the default and positive of

the alternative. The opposite happens in the inconsistent situation. The decision-maker first lists negative aspects of the default and positive of the alternative. Afterwards, the decision-maker lists positive aspects of the default and negative of the alternative. Results show an even more substantial default effect than the previous two experiments. Besides, the inconsistent shows no default effect and no significant difference between choices. I.e. generating a list in the opposite way that is naturally happening elicits the default effect.

This experiment shows the importance of the order of retrieving that information for decisions. Generated lists are from the reference point, that is the default. However, the authors do not explain how these positive or negative aspects are retrieved. This experiment used the example of light bulbs. One of these was better for the environment compared to the other. Recent years there is an increasing awareness of climate change. Using sustainable products is being promoted as one of the solutions. Dinner et al. (2011) did not gather information about personal believes about sustainability. Understandably, it is not easy to extract information. Additional questions beforehand can influence the choice that the participant makes because it gives them ideas what this research is about and can skew their actual choice. Asking afterwards it influence the participant that they give answers to create a view of what they want the researcher to perceive them. Besides, this experiment is to explain the theory and not the real world. Nevertheless, it is not only essential to understand how preferences are formed but also what stimulates these thoughts that, in turn, generates these lists.

Nudges individually are not sufficient, according to Lehner, Mont & Heisenkanen (2016). Nudges should be complementary to information providing. Information generates a list of positive and negative aspects. Dinner, Johnson, Goldstein & Liu (2011) shows that query theory guides the choice. Understanding nudges means to understand how these lists generate. If information providing is a complementary tool, why can it not be that something else is complementary to the nudge? Providing information is changing the environment, the environment where retrievable information is. Nevertheless, if changing this information has an impact, can it be the case that changes in the environment also impact choice? For example, the social environment. A well-formed experiment can confirm this speculation.

Recommendation

Decision-makers weight alternatives as potential losses because the default becomes the reference point. Another reason for the default effect is because decision-makers see it as a recommendation. More precisely, a recommendation from the choice architect or policymaker due to its positive effect. The freedom to choose alternatives makes it a recommendation. Restricting choice is

counterproductive. Restricting makes people go out of their way not to choose the recommendation. Especially when the default is not trusted, or it contradicts their morals or ideas.

The existence of choice architecture does not imply the existence of a choice architect. (Ullmann-Margalit, 1978). Policymakers need to be careful when presenting choices with their respective framework. Default options may be perceived as the option implicitly endorsed by the choice architect (e.g., public policymakers or marketers; McKenzie, Liersch, & Finkelstein 2006). Life consists of many decisions, and not everybody has the time or cognitive strength to think about all of these critically. That is why policymakers have to decide which of the available options to impose on individuals who fail to make a decision (Sunstein, & Thaler, 2003). Helping people will still granting them their freedom is called libertarian paternalism. The critical part is the freedom to choose otherwise. It is the job of the institution to guide decision-makers towards choices that have a positive effect on them. One of the tools available is the default. This nudge creates no restrictions or rules. E.g., With the opt-out policy, there is still the freedom to opt-out of the default. Making the default in a company dinner vegetarian as default, but still offers the freedom to switch to a meat-based dish for the decision-maker. Although not that direct, these defaults carry a message with them. Namely, what the policymaker believes is right.

Default options carry information that can be relevant for the decision-maker. The information from the framework and alternatives can be absorbed differently. McKenzie & Nelson (2003) research the framework of choice when describing a new medicine. Results show the following. When the initial rate is below the new rate, participants choose the framework with survival rate more often compared to the framework that mentions the death rate. This experiment shows the options presented carry information. Participants choose the information providing from the reference point. Besides the fact that the framework contains information from the policymaker, it is also open for interpretation by the decision-maker. How it is decided between two seemingly equivalent description can leak relevant information, which listeners can absorb (McKenzie, 2004). Information leakage seems especially likely in the case of default effects because it is clear to both policymakers and the public that it is effortless to accept the default and effortful to switch from it. (McKenzie, Liersch, Finkelstein, 2006).

McKenzie, Liersch, Finkelstein (2006) study information leakage with the following experiment. In the first experiment, participants answer if they are willing to be organ-donor and if people, in general, should be organ-donor. Above that, what they would set the default if they are policymakers. People who are willing to be organ-donor are more likely to choose the opt-out policy compared to people are who are not, 41% vs 12%. Participants' personal preferences and beliefs about what other people ought to do

were predictive of their chosen default (McKenzie et al. 2006). The results show that what policymakers believe is best will be translated in their choice architecture. The default leaks information. The next experiment shows if the decision-maker absorbs the information.

Two groups have each their default, opt-in or opt-out. 65% of people in the opt-out policy believe policymaker is willing to be organ-donor compared to 2% who believes not. More information leaks with the opt-out policy. 33% believes no information leaks in the opt-out policy compared to 60% in the opt-in policy. All these differences are significant at the 1% level. These results show that decision-makers absorb information through the policy. Policymakers must be careful with their policies because of information leakage and absorption.

Nevertheless, wanted or unwanted, policymakers leak information and decision-makers absorb information. If the cause is implied endorsement, then this suggests that policymakers should either endorse the proposed defaults explicitly or clarify to the public that defaults are not recommendations. (Dinner, Johnson, Goldstein, & Liu, 2011). If the policy is not corresponding with the policymakers, believe the default can be counterproductive. E.g. When policymakers are sceptical about climate change but set policy for sustainability.

McKenzie, Liersch, & Finkelstein (2006) run an experiment based on implicit recommendation, effort, and the default. Participants must read one out of two summaries. The control group has both options available and put on their table and are free to choose. The treatment group only received summary B but was free to choose A. The only requirement is that the participant must ask for it. Results show that 43 out of 44 people in the treatment group choose summary B compared to 17 out of 44 in the control group. After reading and answering questions about the summary participants answer questions on how much they agree with two statements. The first statement is about the implicit recommendation that they read this summary because it appeared that the experimenter wanted them to do this. The treatment group agreed significantly more with this statement than the control group. The second statement is about effort, being that they choose this summary because it costs more effort to choose the other one. Again, the treatment group agree significantly more often with this statement. Implicit recommendations are presumably inferred from defaults in part because it is clear to both policymakers and decision-makers that one option is being made easier to adopt. (McKenzie, Liersch, Finkelstein, 2006).

Although their research shows significant results, it is essential to note that information is extracted directly from the participants. People are not always beware of their behaviour; it can be, for example, an automatic pilot that makes a choice. Explaining this afterwards may be difficult. People do

see the default as an implicit recommendation. Nevertheless, asking them directly can generate answers what researchers want to hear. I.e. participants give desired answers. Therefore, it is essential to extract information indirectly from participants. Besides this, McKenzie, Liersch, Finkelstein (2006) show that default carries information and implicit recommendation. Policymakers leak information about their personal beliefs through policy. Decision-makers absorb this information and sketch an idea about the beliefs of the policymaker.

When studying the default, it cannot be the case not to mention retirement savings. Multiple papers have shown the effectiveness of changing this default. However, McKenzie, Lierch, & Finkelstein (2006) investigate what information lakes and absorbs through these policies. The set-up of the experiment is similar to that of the first one. Two groups with each their default. The first group with the default opt-out in the savings program and the other group opt-in. Participants answer what their view is on the policy decision. 89% of people in the opt-out think that the employee should enrol in the program compared to 6% in the opt-in. 80% in the opt-in group thinks that the employee wants to enrol compared to 11% in the opt-in. 57% in the opt-out group probably enrolled themselves against the 29% in the opt-in. All these differences are significant at the 5% level. The results show that information leaks through the policy. Afterwards, participants answer questions about success.

Results show that 86% in the opt-out believes that it will lead to more people being enrolled compared to the 3% in the opt-in. 71% in the opt-out believes that it would increase the chances that they would enrol against 14% in the opt-in. All these difference between subjects are significant at the 1% level. Although it is evident to participants that there is a default and partially understand it, still the default effect is observable. The nudge does not have to hide. It can be implemented with the framework and still have the desired effect.

Policymakers are choice architects (see Thaler & Sunstein, 2008). They carry the responsibility of understanding the framework of choice. Because as is shown, the framework carries information with itself. Information is un turn absorbed. Lehner, Mont & Heiskanen (2016) write that the implicit recommendation from the policymaker should go hand in hand with information providing. Query theory tells that first, the positive aspects will be listed and afterwards, the negative. Information providing increases the pool of positive or negative aspects. Literature suggests that the default is only effective where preferences are not clear. However, even preferences can change. Nudging preferences and afterwards providing a default what is in line with their preferences is only better for the decision-maker. It shows that the default is not a singular measure to use in the hope that people will stick with it. The

decision-maker must assess the default and believes that it is indeed the best choice for them. Information providing helps with this.

It shows an interaction effect between the default and information providing. When such an interaction exists, then there may be more interaction effects. Information providing increases the pool of information with retrievable information. People behave differently when observable or alone, i.e. in groups or individually. This difference opens doors for research with varying social dynamics. The following part shows literature that confirms that social surroundings impact the choice that people make.

In contrast to the previous part where policies set to default, Huh, Vosgerau, & Morewedge (2014) show with their research that the people around us also create recommendations. They call this social default. When a person is deciding between options for which her preferences are not well-formed, observing the choice of another person makes the option chosen by the other person a social default (Huh, Vosgerau & Morewedge, 2014). The experiment is about unfamiliar Korean snack and teas. An assistant or the previous person sets the social default, for the whole methodology and results see Huh et al. (2014). The following parts show why people around us influence our choice.

Social influence is one of the reasons for the social default and consists out of two aspects—namely informational and normative influence. Deutch & Gerard (1995) explain these aspects. Normative influence is the expectation of others. E.g. when deciding how much to contribute to a group, the choice is based on the contribution of others. Although, when information is publicly available high contribution will create a social default of high contribution. This is based on the desire to behave appropriately in a social setting (Cialdini & Goldstein, 2004).

When people are unsure about their preferences, they start mimicking behaviour. We define choice imitation resulting from social default effects choice mimicry because (like behavioural mimicry) it occurs automatically when consumers do not engage in further deliberation (Huh, Vosgerau, Morewedge, 2014). For mimicry, there must be transparency. It creates guidance and social pressure to show behaviour that is in line with the group. Nudging with social default is possible, but it must not be out of proportion. If it is too extreme, it becomes counter-effective. Huh, et al. (2014) run multiple experiments to test the social default. Results show that social default effect occurs in some situations. It occurs when it seems appropriate to adopt, i.e., when there are no clear preferences. When stakes are high social default effect is not observable, i.e., when there is not much is to lose, social default is effective.

People are aware of the presence of the standard default but not necessarily of its impact. In comparison, people are unaware of the presence of the social default. People stick with the social default

unless decision-makers are certain about their preferences or are able or willing to engage in effortful deliberation before choosing (Huh, Vosgerau, Morewedge, 2014). The social default teaches that the social environment the decision-maker is in impacts the choice. Would this also mean that changes in the social environment will lead to changes in behaviour and hence choices? Will decisions in different group compositions lead to different choices? This can and will be tested in a 2x2 designed experiment to investigate if there is a causal relationship. The current literature gives the first sub hypothesis for answering the research question. *"Contribution will be higher in the default group compared to the non-default group."*

Group composition

The first part of the literature search shows that there are overall three reasons why there is a default effect. One of the reasons is that decision-makers see it as a recommendation. This success is dependent on the amount of social capital. That is, the component of human capital that allows members of a given society to trust one another and cooperate in the formation of new groups and associations (Fukuyama, 1995). Social capital is comprised of civic participation and trusting attitude (Keele, 2007). When in a group, social capital is dependent on interpersonal trust. When people are put together in a group, there is also an amount of social capital.

Gruenfeld, Mannix, Williams, & Neale (1996) underline the importance of groups. It is becoming more relevant because groups now do work that is previously done by individuals. Decisions made by a group of diversified expertise will be higher in quality than those made by employees with more homogenous backgrounds or by any one employee who might have access to the same knowledge but whose single organizational perspective would be more limited (Jackson, 1992). The social setting is relevant for a behavioural economist to understand decisions. One of these social settings is group composition. Decision-makers observe the surround and react to that with their choices. The following part shows how different group compositions based on gender impacts choices. Afterwards, literature about the social dynamics of groups.

Gender effect

Early behaviour economist work like Kahnman & Tversky (1979) show that it is not possible to assume full rationality of humans. Their biases and cognitive power limit decision-makers. An easy way to make a distinguishing in a group is dividing by gender. For simplification, we assume that there are two

genders: man and woman. The following literature will show the gender effect in economic experiments, but not all the results tell the same story. There are contradicting results.

Eckel & Grossman (1998) study the gender effect in a double-blind dictator game. The double-anonymous dictator setting removes risk, the possibility of gender-related subject interactions, and the experimenter effect, leaving only underlying selflessness as an explanation for donating money (Eckel & Grossman, 1998). The double-blind makes it possible to control for confounding factors and truly extract the gender effect. In the experiment, there are ten games. In half of these sessions, there are all-male dictators, and the other half consists of all-female dictators. The respondent group consists of both sexes. Results show that on average, women donate twice as much as men. $1.60 compared to $0.82, with a T-statistic of 2.44 and corresponding P-value below 0.01. When there are fewer confounding factors, women are more altruistic than men.

Andreoni & Vesterlunder (2001) find similar results. Here they also research the gender effect in the dictator game. Participants have eight different allocation problems. Every allocation has its endowment token value. There is the possibility to pass these tokens to the receiver. For example, in one of the allocations keeping a token is worth one and passing it to the receiver is worth three. While in another allocation keeping the token is worth three and passing it to the receiver is worth one. Multiple valuations give different scenarios. It makes giving relatively cheap or relatively expensive. When the relative price of giving is low, men pay significantly more than women. 4.18 compared to 3.01. However, when the relative price of giving becomes expensive, women give significantly more than man. 1.32 compared to 0.67. When the price of giving is low, men appear more altruistic, and when the price is high, women are more generous. (Andreoni & Vesterlunder, 2001). This experiment shows that women prefer an equal distribution and men will mostly only think about themselves or only about others. Men's contribution depends on what kind of impact it has on themselves.

Gender effect is not only studied in the dictator game. Brown-Kruse & Hummels (1993) study the gender effect in the public goods game. In this study theory of Gilligan (1993) is used to predict the behaviour of each gender. There are two behavioural traits—namely contextualism and instrumentalism. In contextualism interpersonal relationships are important. Morality is fulfilling obligations and avoidance of hurting others. While in instrumentalism hierarchical relationships are important. Morality here is individual rights. Gilligan (1993) finds a link between contextualism and women. Also, a link between instrumentalism and men. If this is true, then women will care more for equal distribution between

endowments. Men will prefer to keep their endowment because that is their right and gives them a higher individual payoff. The following part will show the methodology of the experiment.

Groups are formed based on sex. Each group distinguishes itself by community or anonymity. The community group interacts with each other to know them. In the anonymous group, there is no interaction. Players decide to contribute their whole endowment of $1 or nothing. There are a total of six rounds. In the fourth round, the multiplier is increased or decreased. Participants are unaware of this change. This multiplier change makes contributing more or less beneficial. A problem that arises with playing multiple rounds is that there is a learning effect. Participants will start to play strategically. However, again, this is interesting to study because of the repeated interaction. This experiment has four treatments: Community, anonymity, high, and low multiplier. The following part will show the results.

While there is no overall difference between anonymous and community treatment, there is the gender effect. ANOVA shows that in the first-round men contribute significantly more than women. F=5.006 X=0.05. The whole sample shows similar results. Men contribute significantly more than women. F=4.00 P=0.05. Besides, the high multiplier causes for higher contributions. F=7.11 with P=0.01. These results show that men contribute more than women, and a high multiplier increases contributions. Although these results show that men contribute a higher amount, it is essential to note that there is only the possibility to contribute all-or-nothing. Above that, the contribution is $1. The relative cost of giving is not high. Men contribute a higher amount, but only when the relative cost of giving is low and there the situation of all-or-nothing.

Brown-Kruse & Hummels (1993) show that in their sample, it is not merely the case that women are more altruistically. Some comments made by participants show contradicting behaviour relative to what they say. For example, one female subject, upon being told that she would receive an equal dividend from the group fund even if she did not contribute to it, replied, ‘But that would be selfish!’ Yet, in four of the six ensuing rounds, she failed to contribute (Brown-Kruse, Hummels, 1992). She made it clear that it would be against her norms of equal distribution. While men made comments that it seems they would act selfish but show contradicting behaviour. One male, commenting on the experiment, said, ‘I would have screwed these guys over if I thought I could have gotten away with it’. However, he contributed in all six rounds. Another male explained his contribution in five of the six rounds by saying, ‘I was only setting them up’. (Brown-Kruse, Hummels, 1992).

Although these comments are anecdotal evidence and carry no scientific evidence, it is, however, interesting to see that these women, for example, want the researchers to perceive them a certain way.

Namely: altruistic and caring. In comparison, men desire to see as strong and egoist. These contradicting statements show that experiments are a necessary evil to extract behaviour. It is not possible to understand behaviour just by asking. Brown-Kruse & Hummels (1992) find more altruistic men compared to women, but the following study finds contradicting results.

Nowell & Tinkler (1994) study cooperative behaviour in the public goods game. More precisely, the difference between all-female, all-male, and mixed groups. The experiment looks as follows. Groups consist of four members: five all-male, six all-female, and six mixed groups. Each person has an endowment of 62 tokens. Each period the participant can invest in two goods. Good A is private, and good B is public that is multiplied by 1.2 tokens and divided between the four members. Each group plays fourteen rounds. The results show that the free-riding problem increased with each period. With each period contribution decreases, meaning less profit overall. However, the all-female group contributed more tokens compared to the all-male and mixed group in all period. But only significantly more in the tenth and thirteenth round at the 10% significance level.

Data is collected to create a regression analysis. The dependent variable is the contribution. The independent variables are, among other things, gender, total group composition in the previous period, and the autoregressive of the participant's contribution in the previous period. For the complete regression, see Nowel & Tinkler (1994). Results show that the equilibrium contribution in the female group is significantly higher compared to males. The all-female groups contribute a higher amount and thus are more cooperative compared to male. The mixed group equilibrium is not significant. Above that, it has around the same value as the all-male. 15.76 versus 15.56, respectively.

These studies have shown that there is still a lot to learn about gender effect. While Brown-Kruse & Hummels (1993) show that men cooperate more and contribute a higher amount, there is a side note to make. This research has only the decision of all-or-nothing. In comparison, the other literature here can choose the value of their contribution (Nowel & Tinkel, 1994; Eckel & Grossman, 1998; Andreoni & Vesterlunder, 2001). When there is the possibility to choose their contribution, women contribute a higher amount compared to men. Besides that, most literature about the gender effect is from around twenty years ago. Behaviour is not constant; it is ever-changing. The percentage of men working in the Netherlands has stayed stable from 2003 to 2019, 75.58% versus 75.76%. While the percentage of women working has increased in those years. 59.36% versus 66.64% (Central bureau for Statistic, 2020). Women are becoming more represented in the workforce. I.e. increasing amount of responsibility. Because behaviour is changing it is crucial to understand that results from the past may not represent the

population of today. Besides, gender is not the only diversity that the population has. The population of the Netherlands consists of 26.5% of non-Dutch citizens (Central Bureau for Statistics, 2020). It is something that is increasing and needs to be studied to understand behaviour in group composition. The following part will show the current literature about the effect of diversity.

Social diversity

Most literature focusses on the gender effect when playing economic games. However, observe any big cities, and it shows that not the only distinguishing by gender; but also nationality, ethnicity, age. All distinguishes us from each other. Still, groups are essential, but how do these groups form? It is essential to understand how groups form to know the effect of changing this group dynamic. Gruenfeld, Mannix, Williams, & Neale (1996) show that groups are becoming more relevant. The following part will elaborate on their study.

Groups form through selection based on similarity, according to Grunfeld, Mannix, Williams, & Naele (1996). Preexisting relationships bring groups together. These groups have positive affect, smooth interaction, and strong commitment but they can also lack the diversity on which they were meant to capitalize, undermining their potential problem-solving effectives (Jackson, 1992). According to the reference, nothing is wrong with forming groups based on preexisting relationships. It makes interaction only more suitable. However, there are some problems with these groups. Namely, they are less likely to share new ideas. The probability that a given piece of information will be mentioned increases with the number of members who are already aware of it (Stasser & Stewart, 1992). Common knowledge is likely more shared within the group. Nevertheless, it also does not mean that diverse groups are superior. These groups have to tackle their problems, e.g. social acceptance within the group. People are less likely to share their ideas or opinions because they still need to be accepted by the group. Above that, cognitive conflicts arise with diversity. A person experiences mental discomfort when someone confronts them with new information that contradicts their prior beliefs and ideas. Grunfeld et al. (1996) research how familiar and unfamiliar groups work together and how they perceive this cooperation. The following part will elaborate on this.

Each group has to solve a murder. There are three types of groups. One where all members are familiar, the other with two familiar and one stranger, and in the last group, only strangers. Each participant receives information with clues individually. Information received is full or partial. The partial information is unique for each participant. Putting every partial information together will generate all the clues. After solving, participants answer questions about how comfortable they were in their group.

Familiar groups felt significantly more comfortable working together, more comfortable expressing disagreement and more effective than groups of strangers (Grunfeld, Mannix, Williams, & Naele 1996). For the whole methodology and all the results between groups, see Grunfeld et al. (1996). The following part will show only a few results.

When there is full information for all participants, stranger groups outperform familiar groups. However, when information is partial, i.e. each group member has their unique information, familiar groups outperform individual groups. Grunfeld, Mannix, Williams, & Naele (1996) show that the strangers' group prefer to make a democratic decision. I.e. each member tells their suspect and chooses the majority choose instead of pooling together their information. Familiar groups have no problem discussing information and thus pooling together their knowledge. Within familiar groups, there is already the social acceptance to share information. Above that, they are feeling comfortable enough to share it, disprove or confirm groups members ideas. Both groups have their advantages. Diverse groups carry alternative views on problems and bring unique solutions. However, the problem that arises is that there is less social comfort to share innovative views. Forcing groups to be diverse will not have the desired effect. It is crucial to make the diverse group also a familiar group. Familiarization creates comfort and social acceptance to share beliefs and ideas.

The previous shows what effect group composition has on problem fixing. Unfortunately, there is not much literature about nationality diversity in group forming. However, there is the work of Habyarimana, Humphreys, Posner, & Weinstein (2007). They study the effect of ethnicity on public goods provision. This research, however, is not about the effect, but about the reason is that there is a difference in public goods provision. The following part shows the methodology of the study.

The experiment runs in Kampala Uganda, where there is extreme diversity. In their experiment, they look for places with inadequate public goods provision and high diversity. Choosing the “participants” can result in selection bias due to omitted variables that are specific to that region. Therefore, carefulness is necessary for observing the results. Three hundred local individuals are gathered and randomly selected. Interacting is in two ways: one where participants are face-to-face with others and easily observes others. The other way is by showing a box with all the players, including themselves. The players must not know each other to control for confounding factors. Participants answer questions if they know the others. This paper refers to the reader Habyarimana, Humphreys, Posner, & Weinstein (2007) for the whole methodology. The following part is about the main conclusion of the study.

Collective action in the public goods game comes from norms that are controlled and eventually policed when people do not cooperate. Social connections bring cooperation between individuals. Shared ethnicity does, however, not bring better effectiveness on joint tasks. These results are contradicting with Gruenfeld, Mannix, Williams, & Naele (1996). Overall, results show that success from homogenous groups comes from reciprocity and punishing deviating behaviour. Similar to the work of Grunfeld, Mannix, Williams, & Naele (1996) shows that familiarity enhances cooperation within groups. Cooperation comes from knowing the other group members and relying on social norms within a group. Diversity cannot be forced onto people. It is the task for the designer to create the feeling for individuals that they form the groups themselves. Diverse groups must find ways to create relationships to form natural groups to obtain the full potential of groups.

Here ends the literature review. The first part shows the default effect and the reasons behind it. Varying group compositions impact behaviour and decisions. Previous literature shows different behaviour of men and women in economic experiments. It shows that women behave more altruistically and men more selfish. This difference means that when people see other men, they will behave more selfish because they believe that others will do the same. However, in a diverse environment, overall well-being is more important. This expectation generates the second sub hypothesis: *"Contribution in the diverse group is higher than in the homogenous group."*

People observe the default as a recommendation. When this default is in a more inclusive group, dynamic people believe that others will adhere to it. However, when presenting in a homogenous setting, people believe that self-satisfaction is more important, and other players will not follow the recommendation. Choice mimicry is higher when there is the belief that others also care about the well-being of the group. Therefore, to answer this suspicion, the third sub hypothesis is as follows: "Contribution *is higher with the default option compared to no default within their respective group composition.".* The next parts are about the experiment used to answer the research question.

Methodology

Experimental design

The following parts show the methodology of the experiment. First, the experiment design and measurement of control variables. After the measurement, the study shows the design of the public goods game. Here ends the experiment design; afterwards, the methodology of the hypothesis testing and analysis of the results.

Trust measurement

A reliable measurement tool is essential for creating a comparable measurement trust. Multiple papers have come up with different kinds of questions create a comparable measurement (Welch, Hinnant & Moon, 2005; Teo, Srivastava & Jiang, 2008; Bélanger, Carter, 2008; Fischbacher, von Rosenbladt, Schupp & Wagner, 2002; Glaeser, Laibson, Scheinkmen, Soutter, 2000). Most of these previous studies use the measurement of governmental trust for the United States government. Minor modification made these questions applicable to the Dutch government. Appendix 1 & 2 shows the questions used for the measurement of governmental trust and past trusting behaviour.

Public goods game

The public goods game investigates if there is a significant difference in contribution due to the treatments. One of these treatments is a diverse group, and the other is the default option. Two treatments give four possible scenarios: Diverse – default, diverse – no default, homogenous – default, and homogenous – no default. Qualtrics randomly assigns each participant to one of the four possible games. This randomization solves for selection bias. The discussion shows the internal validity of the experiment. The results have baseline checks to assure that the samples come from the same population.

This research wants to visualize the social environment for the participant. Therefore, has this research created multiple avatars with the Google Chrome Webstore app “Avatar Creator”. These avatars make it possible to visualize the group composition somehow. The control group, homogenous, have two overall characteristics. Namely, being white and male. The treatment, diverse, consists of multiple ethnicities and multiple genders. Figure 1 & 2 show the group compositions.

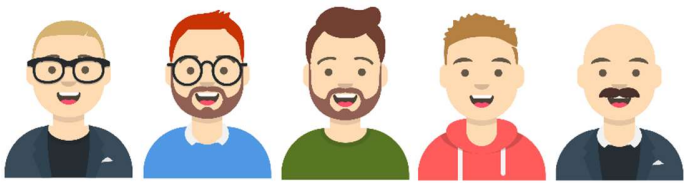

Figure 1. Homogenous group.

Figure 2. Diverse group.

The description of the public goods game show two examples of different contribution. See appendix 3 for the full description. These examples show the participant that it is possible to choose whatever amount of the endowment to contribute.

The default option is contributing the maximum amount and present two out of four possible games. Each in different group composition. The data obtained from the public goods game is on the ratio scale. There is a ranked relationship, meaning higher contribution has a higher ranking. Above that, the difference between numbers have meaning, and there is a true zero point at its origin. After the public goods game, the participants see three hypothetical situations where they have to indicate what they would do.

Each hypothetical scenario is with a default option. The participant answers what they would to in that situation. Answers are to investigate if there is a correlational relationship between the amount of trust extracted and the likelihood of choosing the default, further explanation in the methodology of hypotheses testing. In the first scenario, there is a dinner invitation. The default option of the dinner is non-meat, but there is the freedom to choose an alternative. In the second scenario, the participant just moved and has to choose an energy provider. The default option is staying with buildings collective green energy provider. Again, there is the freedom to choose an alternative. The final scenario the participant receives a company car with a new job. The default option is an electric car with the freedom to choose an alternative car. Default is already selected thanks to Qualtrics.

Lastly, the participant answers questions about their characteristics, age and gender. Another personal characteristic of interest is the annual income of the household where they grew up and their nationality. Education level is not relevant because a big part of the sample will consist out of students. In discussion with the internal validity shows why this is not a problem. Here ends the experiment design. The following part shows how to falsify or confirm the hypotheses.

Hypothesis testing

Statistical tests and regression analysis shows the validity of the hypotheses. However, it is essential to know if the data is suitable. Two kinds of tests are available: parametric and nonparametric. Parametric tests have the following assumptions.

1. Observations are independent.
2. Observations are drawn from a normally distributed population.
3. In the case of two groups, both have the same variance.

4. Variables must be measured in an “interval scale”.

This research cannot assure the second and third assumptions. Therefore, parametric tests are not suitable. Luckily, nonparametric tests have fewer assumptions. For nonparametric tests, only the observations must be independent. For an observation to be independent, someone else’s observation cannot influence theirs. Independence cannot be known for sure but assuming that everybody filled in their survey individually. The significance level indicating if a result is significant is at most 10%. P-values above the 10% significance level give insignificant results.

Nonparametric test

With the first hypothesis, the default option is the treatment. The control group has no default. The dummy *“default”* takes value one when the participant receives the default and zero otherwise. The dummy *“max”* takes value one when the participant contributed the maximum amount, ten, and zero otherwise.

These generated dummies make it possible to run the nonparametric tests. The tests run between subjects and not a theoretical benchmark. Therefore, the Mann-Whitney U-test is suitable. Strictly speaking, this test looks if the distribution of the two samples differs from each other. However, often it is used to look at the differences between the median or average. The Mann-Whitney U-test uses the following null hypothesis:

$$H0: \mu_1 = \mu_2$$
$$Ha: \mu_1 \neq \mu_2$$

With μ_1 being the median of the default group and μ_2 the median of the non-default group. This test will investigate if the median contribution significantly differs between the treatment and control group. Above that, investigating the distribution of people contribute the maximum amount between the treatment and control group. Splitting the groups up by gender makes it clear when there is a difference, which gender causes it. Results show the average and test statistics.

The second hypothesis is when the treatment is the diverse group and the control group the homogenous. New dummy called *“diverse”* takes value one when the participant received the treatment of diverse and zero when they played in the homogenous group. Like the first hypothesis, the Mann-Whitney U test will test the median contribution between subjects. Above that, the distribution of people contributing to the maximum amount.

The results show additional robustness checks. The first robustness check is dividing the sample to create better comparable samples. I.e. between subjects having or not having the treatment of default. The dummy *"dd"* takes value one when the participant played in the diverse – default group and zero when played in the homogenous – default group. The robustness checks use the same methodology as the second hypothesis. I.e. testing the median of contribution and distribution of sticking with the default between subjects. This robustness check shows the effect of the double treatment, or interaction effect, against the control group of homogenous with the only treatment of default. The same is one with the single treatment of diverse versus the control group with no treatment, i.e. diverse – no default versus homogenous – no default.

The third hypothesis is about the treatment of default within their respective group composition. Dummies distinguish treatments within their respective group composition. Taking value one when receiving the treatment of default and zero when not within the diverse group and another dummy for within the homogenous group. First testing the whole sample and afterwards dividing by gender. Mann-Whitney U test generates the test-statistics. This research believes that there is a ranked relationship between the four samples. The Jonckheere test answers these beliefs. The following part elaborates on the Jonckheere test.

This research believes that the amount invested in the diverse group will always be higher in the homogenous group. Above that, previous literature shows that the default option is effective. Previous literature in combination with rational thinking, it can be said that contribution in diverse will be higher than homogenous and default higher than nondefault. Each sample receives a numerical value before the test. Because this research believes that there is a ranking order between samples, what numerical value each sample gets is crucial. The group with the lowest contribution gets value one and highest contribution value four. Giving numerical values creates the following ranking order.

$$\theta_{homogenous-no\ default} < \theta_{homogenous-default} < \theta_{diverse-n\ \ default} < \theta_{diverse-default}$$

With θ being the median, the ranking order is a priori. The Jonckheere test will test the following null hypothesis:

$$\mathrm{H0}: \theta_1 = \theta_2 = \theta_3 = \theta_4$$

$$\mathrm{HA}: \theta_1 \leq \theta_2 \leq \theta_3 \leq \theta_4$$

The Jonckheere test provides robustness and answers if the treatment ranking order is as predicted. Six regressions run additionally to the nonparametric test for robustness. The regression analysis investigates if treatments significantly impact contribution. The next part shows the methodology.

Regression analysis

In all regressions, the dependent variable will be the contribution. When variables of interest are significant, then treatments impact the contribution made, ceteris paribus. The dummies of interest are *diverse, default*, and the interaction term *defaultxdiverse*. The last dummy is also called double treatment. Each dummy will be run once without any control variables and with control variables. These give the following six regressions:

$$Contribution\ (1) = \beta_0 + \beta_1 * default + \varepsilon_i$$

$$Contribution\ (2) = \beta_0 + \beta_1 * default + \beta_i * control_i + \beta_j * control_j + \varepsilon_i$$

$$Contribution\ (3) = \beta_0 + \beta_1 * diverse + \varepsilon_i$$

$$Contribution\ (4) = \beta_0 + \beta_1 * diverse + \beta_i * control_i + \beta_j * control_j + \varepsilon_i$$

$$Contribution\ (5) = \beta_0 + \beta_1 * defaultxdiverse + \varepsilon_i$$

$$Contribution\ (6) = \beta_0 + \beta_1 * defaultxdiverse + \beta_i * control_i + \beta_j * control_j + \varepsilon_i$$

Results

Descriptive statistics

The raw data collects through the survey software Qualtrics. One hundred thirty-five participants played the public goods game. Participants have an average age of twenty-four. In the sample, 49.63% is female, and 50.37% is male. This research investigates the sample for inconsistencies and outliers. Appendix 4 shows the descriptive statistics of the control variables. A single or low observation will disrupt the tests and give wrong statistical results. Therefore, this research drops three observations.

Baseline check

Table 1 is the baseline check from the four samples for continuous and binary variables. Here shows the average of each group. Just looking at averages says nothing. Nonparametric tests are necessary to know if it is possible to continue with these samples. With the Kruskal-Wallis, it is possible to test if the

median of the different samples come from the same population. Kruskal-Wallis uses the following null hypothesis with θ_i being the median of group i:

$$\text{H0: } \theta_1 = \theta_2 = \theta_3 = \theta_4$$

$$\text{Ha: } \theta_i \neq \theta_j for\ some\ groups\ i\ and\ j$$

The last column shows the obtained P-value. Taking a significance level of 5%, if the obtained P-value is above 5%, we fail to reject the null hypothesis. Test results show that the median across all samples does not significantly differ from each other.

Control variable	*Homogenous no default*	*Homogenous default*	*Diverse no default*	*Diverse Default*	*P-value*
Age	25	24.216	23.038	24.853	0.204
Male	0.529	0.514	0.50	0.441	0.896
Who runs the government? *0: Few big interest* *10: Benefit of all the people*	6.324	6.081	6.0	6.412	0.764

Table 1: Baseline checks continuous and binary variables with corresponding P-values with the Kruskal-Wallis test.

Table 2 shows the baseline checks of the categorical variables. These control variables are categorical, and because of that, Kruskal-Wallis is not useable. The Fisher Exact test suits the new requirement. Fisher Exact tests the following null hypothesis:

$$H0: The\ classes\ are\ evenly\ distrubted\ over\ the\ samples$$

$$Ha: The\ samples\ differ\ in\ distrubution.$$

Classes, in this case, are the possible answers that the participant could choose. E.g. how much they trust the government on the Likert-scale. Table 2 shows the P-value obtained for each categorical variable. The mean is left out because it tells nothing about the data due to it being on the ordinal scale. The significance level used is again 5%.

Control variable	P-value
Annual income	0.783
Trust the government	0.880
Wasting taxes	0.259
Act in the interest for me	0.889
Lend money	0.523
Lend possessions	0.355
Leave unlocked.	0.597

Table 2: Baseline checks categorical variables with corresponding P-values with the Fisher Exact test.

All obtained P-values are above the significance level. There is no significant difference between participants for the control variables, i.e., all samples are drawn from the same population. The baseline checks assure this study that the data is useable for further testing. The following part will continue with hypotheses testing.

Hypotheses testing

The dummy *default* takes value one when the participant played with a default option provided and zero otherwise. Table 3 shows the averages of the total sample and divided by gender. Testing the difference of the total sample with the Mann-Whitney U test gives a Z-statistic of -0.773 with P-value of 0.4398. This P-value is above the significance level of 10%, and therefore we fail to reject the null hypothesis. There is no significant difference due to the treatment. Testing the difference for men and women generates both insignificant test results with the Mann-Whitney U test. Z-value of 0.131 with P-value 0.8961 and Z-value of -1.25 with P-value 0.211, respectively.

Treatment	*Total*	*Men*	*Women*
No default	€6.41	€6.71	€6.10
Default	€6.89	€6.65	€7.11

Table 3: Average contribution by gender in the first sub hypothesis. ** p<0.10, ** p<0.05, *** p<0.01.*

Table 4 shows the proportion results of the first sub hypothesis. Testing with the Mann-Whitney U test the total average generates a Z-value of -0.324 with P-value of 0.746. We fail to reject the null hypothesis. There is no statistical difference. The difference for men generates a Z-value of 0.526 with P-value 0.5744. Testing the difference for women gives a Z-value of -1.088 with P-value of 0.2767. All are above the significance level of 10%, and therefore we fail to reject the null hypothesis. Receiving the treatment of default gives no statistical difference. The discussion provides light on what this means for the first sub hypothesis.

Treatment	*Total*	*Men*	*Women*
No default	0.367	0.452	0.276
Default	0.394	0.382	0.405

Table 4: Proportion of participant contributing the maximum amount with the default treatment. ** p<0.10, ** p<0.05, *** p<0.01.*

Table 5 shows the average contribution with the diverse group composition being the treatment. Contribution increases with €0.99 for the total sample. This increase is significant at the 10% level: Z-value is -1.860 with P-value of 0.0628. Testing the difference for men gives a Z-value of -1.526 with P-value 0.1271 and women give Z-value of -1.154 with P-value 0.2484. Both are above the significance level of 10%, and therefore we fail to reject the null hypothesis.

Treatment	*Total*	*Men*	*Women*
Homogenous	€6.22*	€6.12	€6.32
Diverse	€7.21*	€7.41	€7.03

Table 5: Average contribution by gender with diverse treatment. ** p<0.10, ** p<0.05, *** p<0.01.*

Table 6 shows the proportion results of the second sub hypothesis. Although, there is an increase in the mean for the total sample testing shows that this difference is insignificant: Z-value= -0.755 with P-value 0.4503. The same accounts for both genders. Men: Z-value= -1.195 with P-value of 0.2321 and women: Z-value = 0.078 with P-value of 0.938.

Treatment	*Total*	*Men*	*Women*
Homogenous	0.352	0.351	0.353
Diverse	0.417	0.50	0.344

Table 6: Proportion of participant contributing the maximum amount with diverse treatment. ** p<0.10, ** p<0.05, *** p<0.01.*

The dummy *"dd"* takes value one when participant plays in the diverse – default group and zero when in the homogenous – default group. Table 7 shows the average of both samples. Contribution increases with €1.68 when receiving the double treatment. This difference is significant at the 5% level – Z-value = -2.2446 with P-value of 0.0144. We reject the null hypothesis that the median is the same. Receiving the double treatment significantly increases contribution compared to the single treatment of default, ceteris paribus. Also, men increase their contribution with €2.54. Testing this difference generates a Z-value of -2.311 with P-value of 0.0208. This P-value is below the significance level of 5%. However, the difference for women is not significant: Z-value of -1.143 with P-value of 0.2528.

Treatment	*Total*	*Men*	*Women*
Homogenous default	€6.08**	€5.53**	€6.67
Diverse default	€7.76**	€8.07**	€7.53

Table 7: Average contribution by gender with diverse treatment and both default present. ** p<0.10, ** p<0.05, *** p<0.01.*

While table 7 shows the average contribution, table 8 shows the proportion of participants sticking with the default. The Mann-Whitney U test generates a Z-value of -1.734 with P-value of 0.083 testing the difference for the total sample. This P-value is below the significance level of 10%, and therefore we reject the null hypothesis. The proportion of people sticking with the default is significantly higher in the diverse group compared to the homogenous group. The number of men sticking with the default almost triples. Testing this difference generates a Z-value of -2.286 with P-value of 0.0223. This P-value is below the significance level of 5%. There is a statistical difference between the two samples for men. For women, however, the difference is not significant: Z-value of -0.196 with P-value of 0.8443.

Treatment	*Total*	*Men*	*Women*
Homogenous default	0.297*	0.211**	0.389
Diverse default	0.50*	0.60**	0.421

Table 8: Proportion of participant sticking with the default divided by gender with diverse treatment and both default present. ** p<0.10, ** p<0.05, *** p<0.01.*

Previous shows the interaction effect compared to the single treatment of default. Table 9 shows the average contribution comparing homogenous with diverse when there is no default. There is a slight increase in the mean for the total sample. However, this is above the significance level of 10%: Z-value = -0.116 with P-value 0.9078. Testing the difference for and women generate insignificant differences. Men obtain a Z-value of 0.148 with P-value of 0.8824 and women Z-value of -0.359 with P-value 0.719.

Treatment	*Total*	*Men*	*Women*
Homogenous – no default	€6.37	€6.75	€5.94
Diverse – no default	€6.48	€6.65	€6.31

Table 9: Average contribution by gender with diverse treatment and both no default present. ** p<0.10, ** p<0.05, *** p<0.01.*

In all cases, there is a drop in people contributing the maximum amount when receiving the treatment of diversity. Table 10 shows the proportion results. Testing these differences gives the following test statistics. For the total sample: Z-value = -0.755 with P-value 0.4503. Testing generates a Z-value of 0.627 with P-value 0.5309 and women: Z-value = 0.481 with P-value of 0.6304. Table 11 & 12 give a summary of the test results of the second hypothesis.

Treatment	*Total*	*Male*	*Female*
Homogenous – no default	0.412	0.50	0.313
Diverse – no default	0.308	0.385	0.231

Table 10: Proportion of participant sticking with the default divided by gender with diverse treatment and no default present. ** p<0.10, ** p<0.05, *** p<0.01.*

Contribution results	*Total*	*Males*	*Females*
Diverse compared to homogenous	Average = +€1.01 Z-value = -1.86 P-value = 0.0628	Average = +€1.29 Z-value = 1.526 P-value = 0.1271	Average = +€0.71 Z-value = -1.154 P-value = 0.2484
Diverse default compared to homogenous default	Average = +€1.68 Z-value = -2.2446 P-value = 0.0144	Average = +€2.54 Z-value = -2.311 P-value = 0.0208	Average = +€0.86 Z-value = -1.143 P-value = 0.2528
Diverse no default compared to homogenous no default	Average = +€0.11 Z-value = -0.116 P-value = 0.9078	Average = -€0.10 Z-value = -0.148 P-value = 0.8824	Average = +€0.37 Z-value = -0.359 P-value = 0.719

Table 11: Contribution results of the second hypothesis.

Proportion results	*Total*	*Males*	*Females*
Diverse compared to homogenous	Average = +0.065 Z-value = -0.755 P-value = 0.4503	Average = +0.149 Z-value = -1.195 P-value = 0.2321	Average = -0.009 Z-value = 0.078 P-value = 0.938
Diverse default compared to homogenous default	Average = +0.203 Z-value = -1.734 P-value = 0.083	Average = +0.389 Z-value = 2.286 P-value = 0.0223	Average = +0.032 Z-value = 0.196 P-value = 0.8443
Diverse no default compared to homogenous no default	Average = -0.104 Z-value = -0.775 P-value = 0.4503	Average = -0.115 Z-value = 0.627 P-value = 0.5309	Average = -0.082 Z-value = 0.481 P-value = 0.6304

Table 12: Proportion of maximum contribution results of the second hypothesis.

Now looking at the default effect within their respective group composition. Table 13 shows the average contribution and proportion results. There seems an increase in the mean in both cases. However, testing this difference shows that contribution does not significantly differ: Z-value = -1.624 with P-value

of 0.1044. This research barely fails to reject the null hypothesis. Testing the difference for men gives insignificant differences: Z-value -1.211 with P-value 0.2258. The same accounts for women: Z-value = -1.171 with P-value 0.2417. All are above the significance level of 10%, and therefore we fail to reject the null hypothesis.

Similar results show for the proportion. There is a difference in the mean, but testing is essential to know if there is a statistical difference. Testing the total sample gives a Z-value of 1.485 with P-value 0.1376. This P-value is just above the significance level of 10%, and we fail to reject the null hypothesis. The same accounts for the men: Z-value = -1.116 & P-value 0.264; and women: Z-value = -1.096 with P-value 0.2733.

Contribution	*Total*	*Male*	*Female*
Diverse – no default	€6.48	€6.65	€6.31
Diverse – default	€7.76	€8.07	€7.53
Proportion maximum			
Diverse – no default	0.308	0.385	0.231
Diverse – default	0.50	0.60	0.421

Table 13: Average contribution and proportion results of the effect of the default treatment within the diverse group. ** p<0.10, ** p<0.05, *** p<0.01.*

Previous results show the default effect in the diverse group. The following results are from the default within the homogenous group. There is a decrease in contribution when administering the default. However, testing with the Mann-Whitney U test proves insignificant at the 10% level: Z-value 0.324 with P-value 0.7462. The decrease in proportion is also insignificant for the total sample: Z-value 1.002 with P-value 0.3165. The only significant difference comes from the proportion decrease from men when receiving the treatment of diversity: Z-value 1.818 with P-value 0.0696. This is below the significance level of 10%. There is a statistical difference for men.

Contribution	*Total*	*Men*	*Women*
Homogenous – no default	€6.37	€6.75	€5.94
Homogenous – default	€6.08	€5.27	€6.67
Proportion maximum			
Homogenous – no default	0.412	0.50*	0.231
Homogenous – default	0.30	0.211*	0.389

Table 14: Average contribution and proportion results of the effect of the default option within the homogenous group. ** p<0.10, ** p<0.05, *** p<0.01.*

The Jonckheere test tests the ranking order of contribution for the whole sample. Table 15 shows the descriptive statistics of all the public goods games. Running the test generates a test statistic of 1.911 corrected for ties with the corresponding P-value of 0.0233. This P-value of below the significance level of 5% and therefore we reject the null hypothesis. There is an ascending order for contribution between the groups. Here end the nonparametric tests. The following part shows robustness checks in the form of regression analysis.

(Jonckheere) group	*Frequency*	*Percentage*	*Average contribution*
Homogenous – no default	34	29,95%	€6.37
Homogenous – default	37	28,24%	€6.08
Diverse – no default	26	19,85%	€6.48
Diverse – default	34	25,95%	€7.76
Total	131		

Table 15: Distribution among the different group compositions.

Regression analysis

In addition to the hypothesis testing, regressions run to investigate the effect of the dummy treatments. Appendix 5 & 6 show the full regressions. Table 16 show only the treatment dummies and

significant control variables. Regression (1) and (2) show that the dummy *default* has no significant impact on the dependent variable of contribution. I.e. being in the default group has no significant impact on the contribution made.

Regression (3) and (4) show that when being in the diverse group compared to being in the homogenous group significantly increases contribution, ceteris paribus. The highest absolute impact comes from the dummy *defaultxdiverse* in regression (5) and (6). That is the dummy for receiving the double treatment or interaction effect. Of all treatments, this increases contribution the most. The single treatment of default has no significant impact, but when combined with a diverse group, it becomes significantly positive.

A few of the control variables have a significant impact on the independent variable. One of those is the amount of trust that the participant has that the Dutch government does what is right. Because this is a categorical variable that coefficient means relative to someone answering that they trust them some of the time. In all three regressions, (2), (4), and (6), show that a higher amount of trust relative to trusting some of the time significantly increases the contribution made, ceteris paribus.

What was unpredicted is the effect of the categorical variable of wasting taxes. The coefficients shown are all relative to someone answering, *"definitely yes".* Increasing the belief in the government not to waste taxes significantly decreases the contribution made, ceteris paribus.

From all the variables of past trusting behaviour, only the categorical variable of lending possessions shows significant impact. Lending possessions increasingly more often increases the contribution made relative to someone lending possessions never, ceteris paribus. I.e. someone who trust other people more, because they lend out their possessions, increases their contribution. Someone who does not trust other people with their possessions also does not trust other people in this public goods game.

Variables	**Contribution (1)**	**Contribution (2)**	**Contribution (3)**	**Contribution (4)**	**Contribution (5)**	**Contribution (6)**
β (Default)	0.471 (0.555)	0.537 (0.551)				

β (Diverse)			0.990* (0.542)	1.369** (0.542)		
β (DefaultxDiverse)					1.476*** (0.566)	1.717*** (0.557)
β (Trust in to do what is right) Relative to some of the time						
• Most of the time		1.675** (0.758)		1.670 ** (0.729)		1.711 ** (0.738)
• Almost always		2.032** (1.024)		2.002 ** (0.993)		1.992 ** (0.984)
β (Wasting taxes) Relative to definitely yes						
• Probably yes		-3.120 *** (0.948)	-	-3.093 *** (0.979)		-3.332 *** (1.018)
• Might or might not		-3.457 *** (0.985)		-3.337 *** (1.040)		-3.534 *** (1.049)
• Probably not		-3.357 ** (1.034)		-3.245 *** (1.058)		-3.612 ** (1.109)
• Definitely not		-7.828 **** (1.615)		-7.431 **** (1.649)		-7.910 **** (1.741)
β (Lend possessions) Relative never						
• Less than once a month		3.157 **** (0.965)		3.805 **** (1.037)		3.415 *** (0.955)
• At least once a month		2.829 *** (1.057)		3.224 *** (1.106)		2.947 *** (1.043)

• At least once a week		3.085 ** (1.459)		3.255 ** (1.467)		3.017 ** (1.432)

*Table 16: Regression analysis on the contribution. * p<0.10, ** p<0.05, *** p<0.01, **** p<0,0001*

Conclusion

This study tried to answer the research questions: *"What the effect is of group composition on the success rate of the default in the public goods game?".* Three sub hypotheses assist the answer to this question.

The first sub hypothesis looks at the effect of just the default treatment. Although after administering the default in the public goods game show an increase in the mean of contribution and proportion of participants sticking with the default, there is no significant effect. All differences are above the significance level of 10%. Therefore, we reject the first sub hypothesis that contributions will be higher in the default group compared to the non-default group. Receiving the default treatment does not increase contribution.

The second sub hypothesis believes that contribution to the diverse group will be higher compared to the homogenous group. Comparing contribution made in the diverse and homogenous group shows that contribution increases when participants play in the diverse group. The average contribution in the homogenous group is €6.22, while in the diverse group it is €7.21. This difference is significant, P=0.0628<0.10. When both samples have the default present, being in the diverse group significantly increases contribution from €6.08 to €7.76; P=0.0144<0.05. The proportion of people sticking with the default in the same samples significantly increases from 0.297 to 0.50; P=0.083<0.10. These results show the interaction effect of the default and diverse not only for the whole sample but also for men. Men increase their contribution from €5.53 to €8.07; P=0.0208<0.05 and almost tripling the proportion of sticking with the default from 0.211 to 0.60; P=0.0223<0.05. The regression analysis shows that the dummy of diverse and interaction dummy of *defaultxdiverse* significantly increases the contribution, ceteris paribus. These results show that the second sub hypothesis is correct. Contribution to the diverse group is higher than in the homogenous group.

Looking at the default effect within-group composition shows interesting results. Implementing the default in the diverse group increases contribution with at least €1 and proportion increases from 0.308 to 0.50. However, these changes are insignificant at the 10%, and therefore we cannot say that there is a significant difference between the groups. The homogenous group shows the opposite behaviour.

Administering the default decreases the average contribution in the mean from €6.37 to €6.08. Men decrease their contribution from €6.75 to €5.27 while women increase their contribution a little from €5.94 €6.67. The proportion shows similar signs of change. Nevertheless, nonparametric tests show that there is no significant difference. Therefore, the third sub hypothesis is false. The contribution is not higher with the default option within their respective group composition.

While the default by itself shows no significant impact, combining it with a diverse group composition significantly increases the contribution and proportion of people sticking with the default. Diverse group composition increases the success rate of the default option in the public goods game.

Discussion

The following parts first contain the results and what these results mean. After the results, this study shows the limitations and provides future research suggestions.

Testing for the first sub hypothesis shows that implementing the default brings no significant difference for contribution in the public goods game. This research believes that this is due to the fact the diverse and homogenous groups have omitted variables. The results show that participant behaves differently in the diverse and homogenous group and that this causes the insignificance of the default treatment because both group compositions are pooled together.

Splitting up the treatments for the second sub hypothesis shows behaviour that is not as the literature suggests. There is a difference in the mean that women contribute a higher amount compared to men in the homogenous group in all cases, with the exception when there is no default present. This is in line with the literature (Eckel & Grossman, 1998; Nowel & Tinkler, 1994; Andreoni & Vesterlunder, 2001). The same accounts for the proportion of people sticking with the default. However, implementing the diverse group causes a difference in the mean between men and women that men contribute a higher amount and stick more often to the default. This difference is in line with the study of Brown-Kruse & Hummels (1993). These results show that the effect of group composition causes a significant behaviour change for men to the adherence of the default. The most significant difference comes from the interaction effect of group composition and default option. This study shows that the complementary effect of nudges and group composition. Besides, that significant behaviour change does not come from the whole sample, but it has a significant selective effect on certain people. These results confirm that it is crucial to understand the target audience when implementing nudges, as is said by Lehner, Mont, & Heiskanen, (2016).

The third sub hypothesis shows that there is a difference in the mean that men contribute a higher amount compared to women in all cases, with the exception where the default is present in the homogenous group. Although these results show no significance, it is relevant to understand the sign of change. While contribution and proportion increase within the diverse group with the default option, the opposite happens in the homogenous group. The default is counter-effective and causes a decrease in contribution and hence cooperation. This decrease is not expected from literature about the default, (Sunstein, & Thaler, 2003; McKenzie, Lierch, & Finkelstein, 2006; Egebark & Ekström, 2016; Sunstein & Reisch, 2014; Pichert & Katsikopoulos, 2008). However, Lombardini & Lanoski (2013) show that forcing a default on people can be counter-effective. The decrease in contribution comes from men. The proportion of men sticking with the default becomes 1/3rd of the control group. This decrease can be caused by the fact that men believe others think just like them. Forcing men in a similarly minded group creates a dislike. This dislike causes them to go out of their way and not cooperate. This dislike confirms that it is not only essential to know the target audience, but also the social environment of the nudge.

The Jonckheere test adds robustness to the strength of the interaction effect of the two treatments. The ascending order shows that default gives a higher contribution than no default and diverse gives higher contribution than homogenous. The complementary of the two treatments also shows in the regression analysis, where both dummies of diverse and the interaction term of default and diverse shows a significant positive effect on the contribution. That the single treatment of default shows no significance and is in line with the work of Lehner, Mont & Heisenkanen (2016). They show that only the default will not have the desired effect and should be complemented with information providing. Although group composition is not information, it is changing the information about the surroundings of the decision, and it complements the default. Men show counter-effective behaviour when presenting the default, but changing the environment of the decision shows adherence to the nudge. The robustness checks confirm Lombardini & Lanoski (2013) that simply forcing a default will have counter-effective results. There needs to be a better understanding of the environment that surrounds the decision-maker.

For an experiment to be internally valid, four of the five precepts of Smith (1982) have to be satisfied: nonsatiation, saliency, dominance, and privacy. For nonsatiation given a costless choice between alternatives, one always be chosen over the other. The derivative is above zero; having more is better. As this research uses contribution, hence money, more money for the participant is better. For salience, we cannot cheat the subjects. Good decisions in the experiment get rewarded, and bad decisions punished. Because this research does not use any financial incentives, this precept is automatically satisfied. The third precept is about dominance. I.e. rewards must be higher than the subjective cost of participating.

Assuming that people are happy to participate because it helps the author gives some reward. Besides, the survey is made a short as possible to lower the subjective cost. Lastly, there is the precept of privacy. The results are secret, and the participant does not know about the choices of other participants. Subjects cannot attach weight on the earnings of others. Satisfying all these precept means that we assure that the results are internally valid.

Selection bias can be a problem because of a good part of the participants are students. However, this should not be a problem for the internal validity of the experiment if there is randomization. The baseline checks give the reassurance that all samples come from the same population. The importance of this comes from the last precept of Smith (1982): parallelism. I.e. given the conditions are the same, the results inside the lab give the same results as outside the lab. The goal of this research is not the describe the real world but to explain the effect of group composition on the default option. Experiment describes how characteristics, or in this case, social environment, influence economic behaviour. This experiment is, therefore, internally valid and is reproducible with the same conditions to obtain similar results.

Due to the Covid-19 outbreak, it was challenging to gather a larger sample size to obtain bigger N-terms for each of the groups. The low N makes it more challenging to identify the significant difference of coefficients. Some insignificant coefficients were expected to be significant. Treatments can cause differences in the means, but the low N-term makes it difficult to find significant relationships. Larger sample size creates better normal distribution that is not disrupted by a few outliers. Another limitation is that participants play with avatars instead of real humans. Lack of resources made it impossible to obtain real people for the public goods game.

For future research, this study suggests studying the effect of group dynamic on contribution to mitigate costs for a collective problem. This future study can investigate how cooperation is between different group dynamics and which group dynamic cooperates better. This study can show how problem solutions work in different group compositions.

This study shows that the single treatment of default has no significant impact in the public goods game. However, making the default complementary with diverse group compositions, it becomes economically powerful and significant. Men adhere more to the default compared to women, but only in combination with diverse group composition. Behavioural economists should study the surroundings of nudges further to understand the effect. Knowledge about the nudge as well as the surroundings helps policymakers to create better well-being.

Bibliography

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293-312.

Bélanger, F., & Carter, L. (2008). Trust and risk in e-government adoption. *The Journal of Strategic Information Systems*, *17*(2), 165-176.

Brown-Kruse, J., & Hummels, D. (1993). Gender effects in laboratory public goods contribution: Do individuals put their money where their mouth is?. *Journal of Economic Behavior & Organization*, *22*(3), 255-267.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591-621.

Central Bureau for Statistics. (2020, July 23). Statistics Netherlands: Population; age, migration background, gender and region [Data file]. Retrieved from https://opendata.cbs.nl/statline/?dl=44A28#/CBS/nl/dataset/82309NED/table

Central Bureau for Statistics. (2020, November 13). Statistics Netherlands: Labor participation [Data file]. Retrieved from https://opendata.cbs.nl/statline/?dl=44A28#/CBS/nl/dataset/82309NED/table

Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: why people choose not to choose. *Journal of Experimental Psychology: Applied*, *17*(4), 332.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The economic journal*, *108*(448), 726-735.

Egebark, J., & Ekström, M. (2016). Can indifference make the world greener?. *Journal of Environmental Economics and Management*, *76*, 1-13.

Evans, A. M., Dillon, K. D., Goldin, G., & Krueger, J. I. (2011). Trust and self-control: The moderating role of the default. *Judgment & Decision Making*, *6*(7).

Freedman, M. R., & Brochado, C. (2010). Reducing portion size reduces food intake and plate waste. *Obesity*, *18*(9), 1864-1866.

Fukuyama, F. (1995). Social capital and the global economy. *Foreign Aff.*, *74*, 89.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The quarterly journal of economics*, *115*(3), 811-846.

Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Goldberg, M. E., & Gunasti, K. (2007). Creating an environment in which youths are encouraged to eat a healthier diet. *Journal of Public Policy & Marketing*, *26*(2), 162-181.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational behavior and human decision processes*, *67*(1), 1-15

Habyarimana, J., Humphreys, M., Posner, D. N., & Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision?. *American Political Science Review*, *101*(4), 709-725.

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2014). Social defaults: Observed choices become choice defaults. *Journal of Consumer Research*, *41*(3), 746-760.

Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging–A promising tool for sustainable consumption behaviour?. *Journal of Cleaner Production*, *134*, 166-177.

Lombardini, C., & Lankoski, L. (2013). Forced choice restriction in promoting sustainable food consumption: Intended and unintended effects of the mandatory vegetarian day in Helsinki schools. *Journal of consumer policy*, *36*(2), 159-178.

Jackson, S. E. (1992). Consequences of group composition for the interpersonal dynamics of strategic issue processing. *Advances in strategic management*, *8*(3), 345-382.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science, 302*(5649), 1138-1339

Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: a query theory of value construction. *Journal of experimental psychology: Learning, memory, and cognition*, *33*(3), 461.

Johnson, E. J., Payne, J. W., & Bettman, J. R. (1988). Information displays and preference reversals. *Organizational behavior and human decision processes*, *42*(1), 1-21.

Kahneman, Daniel; Tversky, Amos (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica*. *47*(2): 263–291

Kallbekken, S., & Sælen, H. (2013). 'Nudging'hotel guests to reduce food waste as a win–win environmental measure. *Economics Letters*, *119*(3), 325-327.

McKenzie, C.R.M. (2004). Framing effects in inference tasks—and why they are normatively defensible. Memory & Cognition, 32, 874–885.

McKenzie, C. R., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, *17*(5), 414-420.

McKenzie, C. R.M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic bulletin & review*, *10*(3), 596-602.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research, 13*(3), 334-359.

Park, C. W., Jun, S. Y., & MacInnis, D. J. (2000). Choosing what I want versus rejecting what I do not want: An application of decision framing to product option choice decisions. *Journal of Marketing Research*, *37*(2), 187-202.

Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, *28*(1), 63-73.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review, 72*(5), 923-955.

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of personality and social psychology*, *63*(3), 426.

Sunstein, C. R., & Reisch, L. A. (2014). Automatically green: Behavioral economics and environmental protection. *Harv. Envtl. L. Rev.*, *38*, 127.

Teo, T. S., Srivastava, S. C., & Jiang, L. (2008). Trust and electronic government success: An empirical study. *Journal of management information systems, 25*(3), 99-132.

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American economic review*, *93*(2), 175-179.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Ullmann-Margalit, E. (1978). Invisible-hand explanations. *Synthese*, *39*(2), 263-291.

Welch, E. W., Hinnant, C. C., & Moon, M. J. (2005). Linking citizen satisfaction with e-government and trust in government. *Journal of public administration research and theory, 15*(3), 371-391.

Appendix

1. Questions governmental trust

How often do you think you can trust the Dutch government to do what is right?

- ◯ Never.
- ◯ Some of the time.
- ◯ Most of the time.
- ◯ Almost always.

Would you say that the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?

Few big interests										Benefit of all the people.
0	1	2	3	4	5	6	7	8	9	10
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

Do you think that people in the government waste a lot of money we pay in taxes?

- ◯ Definitely yes.
- ◯ Probably yes.
- ◯ Might or might not.
- ◯ Probably not.
- ◯ Definitely not.

How much do you agree with the following statement?

"When dealing with a government agency I can trust them that they have the best interest for me."

- ◯ None at all.
- ◯ A little.
- ◯ A moderate amount.
- ◯ A lot.
- ◯ A great deal.

2. Questions past trusting behaviour

How often do you lend money to your friends?

- ◯ Never.
- ◯ Less than once a month.
- ◯ At least once a month.
- ◯ At least once a week.
- ◯ Almost everyday.

How often do you lend personal possessions to your friends? For example, clothes, bicycle, games etc.

- ◯ Never.
- ◯ Less than once a month.
- ◯ At least once a month.
- ◯ At least once a week.
- ◯ Almost everyday.

How often do you intentionally leave your room or house unlocked when nobody is home?

- ◯ Never.
- ◯ Less than once a month.
- ◯ At least once a month.
- ◯ At least once a week.
- ◯ Almost everyday.

3. Description of public goods game.

You are invited to play the public goods game. The game goes as follows:
You are grouped with other people. Every person receives a certain starting amount of money.
Everybody makes the decision how much they contribute to the communal collection. You are free to choose how much you would contribute. You keep the money that you don't contribute. Each contribution is made anonymous, it will be unknown who paid what.
Afterwards, the amount collected will be doubled and divided equally between group members.
End of the game.

Example 1: Everybody in your group of six people starts with an endowment of €2. Every player contributes €2. The total amount collected will be €12. This will be doubled and distributed equally among all players.
Every player receives €24 / €6 = €4.

Example 2: Everybody in your group of six people starts with an endowment of €2. Every player contributes €1. The total amount collected will be €6. This will be doubled and distributed equally among all players. Every player receives €12 / 6 = €2 plus the amount (€1) not contributed.

Now for real.

4. Descriptive statistics trust control variables.

How often do you think you can trust the Dutch government to do what is right?	Frequency	Percentage
Never	1	0.79%
Some of the time	28	20.74%
Most of the time	85	62.96%
Almost always	21	15.56%
Total	135	100%
Would you say that the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?	Average	
Scale one till ten	6.15	
Do you think that people in the government waste a lot of money we pay in taxes?	Frequency	Percentage
Definitely yes	11	8.21%
Probably yes	42	31.34%
Might or might not	36	26.87%
Probably not	39	29.10%
Definitely not	6	4.48%
Total	134	100%
How much do you agree with the following statement?	Frequency	Percentage

"When dealing with a government agency I can trust them that they have the best interest for me."		
None at all	1	0.75%
A little	17	12.69%
A moderate amount	70	52.24%
A lot	36	26.87%
A great deal	10	7.46%
Total	134	7.46%

How often do you lend money to your friends?	Frequency	Percentage
Never	27	20.30%
Less than once a month	67	50.38%
At least once a month	31	23.31%
At least once a week	8	6.02%
How often do you lend personal possessions to your friends? For example, clothes, bicycle, games etc.	Frequency	Percentage
Never	13	9.77%
Less than once a month	56	42.11%
At least once a month	45	33.83%
At least once a week	18	13.53%
Almost everyday	1	0.75%

How often do you intentionally leave your room or house unlocked when nobody is home?	Frequency	Percentage
Never	80	60.61%
Less than once a month	8	6.06%
At least once a month	8	6.06%
At least once a week	14	10.61%
Almost everyday	22	16.67%

5. Regression analysis default

	(1) contribution	(2) contribution
β default	0.471	0.537
	(0.555)	(0.551)
β (Trust)		
Relative to some of the time		0
		(.)
Most of the time		1.675*
		(0.758)
Almost always		2.032*
		(1.024)
β Runs		0.142
		(0.155)
β (Wasting taxes)		

Relative to definitely yes	0
	(.)
Probably yes	-3.119**
	(0.948)
Might or might not	-3.457***
	(0.985)
Probably not	-3.357**
	(1.034)
Definitely not	-7.823***
	(1.615)
β (Lending money)	
Relative to never	0
	(.)
Less than once a month	-0.917
	(0.757)

At least once a month	0.0577
	(0.880)
At least once a week	0.309
	(1.742)
β (Lend possessions)	
Relative to Never	0
	(.)
Less than once a month	3.157**
	(0.965)
At least once a month	2.828**
	(1.057)
At least once a week	3.085*
	(1.459)
β (Leave unlocked)	
Relative to Never	0

	(.)
Less than once a month	-1.347
	(1.484)
At least once a month	-1.359
	(1.158)
At least once a week	2.379*
	(0.911)
Almost everyday	-0.700
	(0.841)
β Annual income	
Relative to below the €30.000	0
	(.)
Between €30.000 & €50.000	0.0750
	(0.885)

Between €50.000 & €100.000		0.734
		(0.861)
Above the €100.000		-0.0959
		(1.006)
β (Male)		-0.661
		(0.586)
β (Age)		0.0641
		(0.0511)
Constant	6.417***	3.701
	(0.424)	(2.333)
Observations	131	130

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6. Regression analysis diverse

	(1) contribution	(2) contribution
β Diverse	0.990	1.369*
	(0.542)	(0.543)
β (Trust)		
Relative to some of the time		0
		(.)
Most of the time		1.670*
		(0.729)
Almost always		2.002*
		(0.993)
β Runs		0.132
		(0.159)
β (Wasting taxes)		
Relative to definitely yes		0

	(.)
Probably yes	-3.093**
	(0.979)
Might or might not	-3.337**
	(1.040)
Probably not	-3.245**
	(1.058)
Definitely not	-7.431***
	(1.700)
β (Lending money)	
Relative to never	0
	(.)
Less than once a month	-1.278
	(0.733)

At least once a month	-0.244
	(0.836)
At least once a week	0.729
	(1.649)
β (Lend possessions)	
Relative to Never	0
	(.)
Less than once a month	3.805***
	(1.037)
At least once a month	3.224**
	(1.106)
At least once a week	3.255*
	(1.467)
β (Leave unlocked)	
Relative to Never	0
	(.)

Less than once a month	-1.267
	(1.426)
At least once a month	-0.732
	(1.129)
At least once a week	2.616**
	(0.875)
Almost everyday	-0.845
	(0.804)
β Annual income	
Relative to below the €30.000	0
	(.)
Between €30.000 & €50.000	-0.0379
	(0.860)

Between €50.000 & €100.000		0.567
		(0.855)
Above the €100.000		0.0349
		(0.985)
β (Male)		-0.659
		(0.576)
β (Age)		0.0677
		(0.0541)
Constant	6.218***	3.070
	(0.384)	(2.379)
Observations	131	130

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7. Regression analysis interaction defaultxdiverse

	(1) contribution	(2) contribution
β DefaultxDiverse	1.476*	1.717**
	(0.566)	(0.557)
β (Trust)		
Relative to some of the time		0
		(.)
Most of the time		1.711*
		(0.738)
Almost always		1.992*
		(0.984)
β Runs		0.132
		(0.153)
β (Wasting taxes)		
Relative to definitely yes		0

	(.)
Probably yes	-3.332**
	(1.018)
Might or might not	-3.534**
	(1.049)
Probably not	-3.612**
	(1.109)
Definitely not	-7.910***
	(1.741)
β (Lending money)	
Relative to never	0
	(.)
Less than once a month	-1.095
	(0.729)

At least once a month	-0.00717
	(0.834)
At least once a week	0.525
	(1.654)
β (Lend possessions)	
Relative to Never	0
	(.)
Less than once a month	3.415***
	(0.955)
At least once a month	2.947**
	(1.043)
At least once a week	3.017*
	(1.432)
β (Leave unlocked)	
Relative to Never	0
	(.)

Less than once a month	-1.591
	(1.469)
At least once a month	-0.899
	(1.114)
At least once a week	2.501**
	(0.891)
Almost everyday	-0.759
	(0.784)
β Annual income	
Relative to below the €30.000	0
	(.)
Between €30.000 & €50.000	0.0555
	(0.882)

Between €50.000 & €100.000		0.563
		(0.863)
Above the €100.000		-0.238
		(0.951)
β (Male)		-0.551
		(0.571)
β (Age)		0.0556
		(0.0514)
Constant	6.289***	3.923
	(0.326)	(2.287)
Observations	131	130

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$