



A fine is a price. Or is it?

A field experiment: how descriptive norms can increase
rule compliance when fines are in place

5-8-2016

Name: Erik Van de Calseijde
Student number: 425264ec
MSc Economics and Business: Behavioural Economics specialisation
Supervisor: Paul van Bruggen

Abstract

Much of the existing deterrence research showed that increasing penalty severity does often not increase compliance. The deterrence hypothesis and the effectiveness of penalties were cast into doubt altogether by one particular study, which argued that “a fine is a price” (Gneezy & Rustichini, 2000, p. 14). Descriptive norms, on the other hand, have effectively increased rule compliance in a wide variety of cases but they sometimes backfire, and it is difficult to predict its effectiveness *a priori*. Hence, tools for policymakers to increase compliance seem rather limited.

Therefore, a natural field experiment was conducted at Erasmus University Rotterdam’s (EUR), aiming to broaden policymakers’ tools for increasing compliance. The descriptive norms treatment entailed presenting late borrowers with the borrowing compliance among the EUR library’s borrowers. The treatment was based both on literature and a preliminary survey and it was hypothesized that it would increase compliance. This means that late borrowers would hand in their overdue books faster if presented with descriptive norms than when not. In order to control for policy changes, the data also allowed for conducting a natural experiment, which consisted of the EUR library’s change in borrowing policy. It was hypothesized that eliminating fines for being one and two weeks due would decrease compliance proportions.

Based on the results three main conclusions were drawn. Firstly, borrowing compliance increased significantly when two, lower fines were eliminated, which might have been affected by a higher fine at a later stage. Secondly, borrowing compliance further increased when borrowers were presented with descriptive norms. Finally, the effectiveness of the norm suggests that an online survey is a practical way to implement norms, since it can predict a norm’s (in)effectiveness *a priori*. These results have relevant implications for policy measures and make a compelling case for further research on the (in)effectiveness of penalties as well as new potential applications of descriptive norms.

Table of Contents

Abstract.....	1
1. Introduction	4
2. Literature Review	5
2.1 Descriptive norms	5
2.2 Fines.....	9
3. Hypotheses Development.....	10
3.1 EUR library's current communication flow.....	11
3.2 EUR library's communication flow in 2015.....	13
3.3 The book compliance proportion	13
3.4 Hypotheses.....	15
4. Experiment	15
4.1 Penalty treatment	17
4.2 Norm treatment	17
5. Results of Study.....	20
5.1 Penalty treatment	21
5.2 Norm treatment	21
5.3 Norm treatment vs penalty treatment.....	22
6. Discussion	24
6.1 Implications.....	24
6.2 Limitations	26
6.3 Recommendations for Future Research	27
7. Conclusions	28
8. Bibliography.....	29
9. Appendices	33

List of Tables

Table 1: University Library's borrowing procedure.....	11
Table 2: Survey statistics of the two-sided Wilcoxon signed-rank test	19
Table 3: P-values Pearson's Chi-squared test and Fisher's exact test	21
Table 4: P-values Pearson's Chi-squared test and Fisher's exact test	22
Table 5: P-values Pearson's Chi-squared test and Fisher's exact test	23
Table 6: P-values Pearson's Chi-squared test and Fisher's exact test	24

List of Figures

Figure 1: Weekly recall flow	12
Figure 2: Monthly number of 1st, 2nd, 3rd and 4th reminders in 2015	12

1. Introduction

Some state that punctuality is a virtue or even “the politeness of kings”, which was attributed to Louis XVIII (Speake, 2015). The Erasmus University Rotterdam (EUR) library’s data, however, clearly suggests that its borrowers disagree about the importance of punctuality. Lending out thousands of books a year, the library is unfortunately also forced to send out thousands of reminders to individuals and institutions that possess overdue books. The problem of overdue books limits the library’s ability to provide its borrowers with books they desire, which can cause frustration among its borrowers. When one does not return a book, this poses a negative externality on other borrowers. It also demands time from the university library’s staff, which in turn restrains time that could be spent on further improving the library’s services.

By sending out recall messages, the EUR library attempts to encourage its borrowers to return the books as soon as possible, after they have missed the (first) return deadline. The library does not reward students, university staff and external borrowers for returning their books on time. Apart from the fact that this would be too costly for the library, timely return of books is also perceived as something one ought to do anyway, with or without any rewards to be gained. The EUR library does have penalties for its borrowers who do not return books on time after being reminded repeatedly for a long time already. However, since January 2016, the library has eliminated the fines that were incurred in the past for the second and third reminder and replaced the fine for the fourth reminder with a higher one. Nevertheless, the problem of overdue books still persists.

Apesteguia, Funk and Iriberry (2013) conducted a randomized field experiment in the public libraries of Barcelona to test how compliance could be increased by sending reminders before the due date. One effective way to increase rule compliance was sending the following general reminder: “If at some point you borrow an item from the library, please remember that you have to return it on time” (Apesteguia, Funk, & Iriberry, 2013, p. 270). In the field of behavioral economics such a priori reminders have also been proven effective in various other settings such as promoting college enrolments (Castleman & Page, 2015), endorsing saving (Karlan, McConnell, Mullainathan, & Zinman, 2016) and collecting delinquent fines (Haynes, Greeb, Gallagher, John, & Torgerson, 2013).

However, the EUR library only sends out reminders once books are due already. In behavioral economics, descriptive norms have been used to alter behavior and increase rule compliance in various environments. Given the scope of descriptive norms, my research question is as follows: can descriptive norms help to increase borrowing compliance at the EUR library? This was tested by means of conducting an online survey and a natural field experiment at the EUR library. In order to control for policy changes, the data allowed for testing the effectiveness of removing two fines and comparing this to impact of descriptive norms.

In Section 2, I review existing literature on descriptive norms and its applications, as well as literature on penalties. Section 3 contains the calculation of the book compliance proportion and my hypotheses, while Section 4 describes the survey and the experiment itself. I present my results in Section 5 and my discussion consists of the implications and the limitations of this study, as well as recommendations for future research in Section 6. Finally, in Section 7, I draw conclusions on the results.

2. Literature Review

2.1 Descriptive norms

In the last decade the field of behavioral economics has grown tremendously and received more and more attention of the public, not least due to bestsellers such as *Nudge: Improving Decisions about Health, Wealth and Happiness* by Thaler and Sunstein (2008). The exact definition of a nudge “is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler & Sunstein, 2008, p. 6). Various countries have embraced behavioral economics and set up centralized ‘nudge units’ in order to look for more opportunities to apply them. Arguably the most famous nudge unit is the Behavioural Insights Team (BIT), which originally started in the UK Cabinet’s Office in 2010 but spun out in February 2014.

Recently the Behavioural Insights Team has promoted descriptive norms as effective means to encourage certain behavior. In fact, the BIT has included “making it social” in its EAST model (Easy, Attractive, Timely & Social) as one of the four principles for applying behavioral insights and it entails “showing that most people perform the desired behavior” (The Behavioural Insights Team, 2014, p. 28). This matches Cialdini, Reno, and Kallgren’s (1990) definition of descriptive norms, which describe the actual prevalence of non-compliance (what people do), as opposed to injunctive norms, which describe the social acceptability of non-compliance (what people approve). In this renowned study on littering in public places, Cialdini et al. (1990) also revealed that descriptive norms had an effect because individuals littered more in littered environments than in clean environments. Furthermore, the size of the descriptive norm effect increased strongly when descriptive norms were made more salient (Cialdini et al., 1990). Consequently, in littered environments individuals actually littered even more when descriptive norms were made more salient (because a confederate littered) than when they were not, while the opposite happened in clean environments (even though both results were nonsignificant) (Cialdini et al., 1990). Although it remains unclear whether descriptive norms are more likely to change behavior than injunctive norms (Berkowitz, 2004), the BIT still specifically promotes descriptive norms and these are therefore also the focus of this study.

By showing what people actually do, descriptive norms are presented as an attempt to nudge individuals to join the status quo. The reasoning is that by showing that most people actually perform the desired behavior, individuals can be encouraged to do the same. One particular phenomenon that showcases this is the bandwagon effect. In 1950, Leibenstein defined the bandwagon effect as "the extent to which the demand for a commodity is increased due to the fact that others are also consuming the same commodity" (Leibenstein, 1950, p. 189). As more 'people hop on the bandwagon', the uptake of products, ideas or beliefs further increases because of the growing adoption by others, regardless of the underlying evidence. Fashion is a clear example where this effect can play a role when demand increases simply because more consumers purchase a good (Long, Fox, & York, 2007). Since this effect can turn preliminary polls, which express the public opinion, into self-fulfilling prophecies (Rothschild & Malhotra, 2014) some countries have banned polls near election time for this reason (Mehravian, 1998).

Regardless of the previously mentioned examples, people tend to underestimate how much their own behavior is influenced by others' behavior (Nolan, Cialdini, Goldstein, & Griskevicius, 2008). But oftentimes people take their peers' behavior as a benchmark of what desirable behavior is and act accordingly (Dolan, Hallsworth, Halpern, & King, 2010). Descriptive norms indicate how that certain behavior by unfolding how prevalent that behavior is. The next section highlights examples where descriptive norms were applied successfully to alter human behavior.

2.1.1 Descriptive norms' successes

One of the BIT's biggest breakthroughs was the tax trial it ran with Her Majesty's Revenue and Customs department, which is responsible for collection of taxes in the UK. When they sent out different social normative messages to taxpayers it resulted in collecting an additional £210 million in tax revenues (The Behavioural Insights Team, 2015). The most effective treatment contained the following descriptive norm: "Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet" (Hallsworth, List, Metcalfe, & Vlaev, 2014). Unlike the study by Blumenthal et al. (2001), this study did specify how many taxpayers pay their taxes on time by revealing a proportion (9/10). The results indicated that descriptive norms had a significantly larger effect as compared to injunctive norms on increasing compliance (Hallsworth et al., 2014), which helps explain the BIT's focus on descriptive norms.

The BIT replicated the use of social normative messages in cooperation with the World Bank in order to increase tax compliance in Guatemala. This time the best performing treatment was the letter that addressed deliberate choice: "Previously we have considered your failure to declare an oversight. However, if you don't declare now we will consider it an active choice and you may therefore be audited and could face the procedure established by law" (Kettle, Hernandez, Ruda, & Sanders, 2014). But the second best performing treatment again addressed descriptive norms: "According to our records, 64.5%

of Guatemalans declared their income tax for the year 2013 on time. You are part of the minority of Guatemalans who are yet to declare for this tax” (Kettle et al., 2014). Even though the letter addressing deliberate choice increased tax compliance the most (1.7 percentage points) the letter addressing social norms was still very effective and increased the rate of payment by 1.4 percentage points. Furthermore, it is conceivable that the effect of social norms could have been stronger if non-compliance was less prevalent in Guatemala and the mentioned compliance rate (64.5%) was as high as in the UK (>90%).

Apart from experiments conducted by the BIT and the aforementioned study on littering the environment by Cialdini et al. (1990), there are other situations in which descriptive norms can affect behavior. For example, Köbis, van Prooijen, Righetti, & Van Lange (2015) provided empirical support for the impact that descriptive norms had on corrupt behavior. Informing participants prior to a bidding game that corrupt behavior was not prevalent reduced the level of corrupt behavior (Köbis et al., 2015). Another study found that comparing a household’s energy use to an efficient neighbor reduces their energy usage by 2% to 4% (Ayres, Raseman, & Shih, 2012). Yet, it is very difficult to predict upfront what the effect of social norms will be exactly, as will be highlighted in the next two paragraphs.

2.1.2 Disputed effects of descriptive norms

Towel reuse is one clear example where the effects of descriptive norms are mixed. The first study on this topic employed descriptive normative messages as follows: “the majority of guests reuse their towels” (Goldstein, Cialdini, & Griskevicius, 2008, p. 472). Their study showed that descriptive normative messages were more effective as compared to appealing to environmental protection (Goldstein et al., 2008). The authors also found that individuals were influenced more when social norms stated that 75% of previous users of that exact same room (provincial norms) had reused their towels, as compared to 75% of this hotel’s guests in general (global norms). According to Goldstein et al. (2008) this implies that the reference group that was physically most proximate lead to the highest participation rates.

However, when Bohner & Schlüter (2014) replicated this field experiment in hotels in Germany they found different results. Even though both descriptive norm messages and standard messages increased towel reuse rates compared to showing no message, descriptive norms were not more effective than the standard message (Bohner & Schlüter, 2014). On the contrary, the standard environmental message was more effective than the descriptive norm messages. Furthermore, the effects of proximity were inconsistent and therefore deviated from what Goldstein et al. (2008) found. Bohner and Schlüter (2014) argue that different levels of knowledge and involvement with an issue like pro-environmental behavior as well as different probabilities of the target behavior might be two factors that affect the impact of applying social norms and that these could differ per culture.

Results from a third study on towel reuse in hotels in Switzerland and Austria by Reese and Steffgen (2014) deviate both from the findings by Bohner and Schlüter (2014) and the findings by Goldstein et

al. (2008). On the one hand, Reese and Steffgen (2014) found that provincial normative messages (75% of guests in this room usually use their towels more than once) lead to a significant increase in towel reuse, as compared to standard environmental messages and compared to global normative messages. On the other hand, there was no significant difference between showing global normative messages and standard environmental messages (Reese, Loew, & Steffgen, 2014).

2.1.3 Backfiring descriptive norms

The predictive power of descriptive norms has been disputed by studies in different contexts, too. One example is a field experiment in Austria that tested the effectiveness of different letters to potential evaders of TV-license fees. The authors showed that messages with descriptive norms, mentioning a 94% compliance rate, proved to have no effect on changing behavior (Fellner, Sausgruber, & Traxler, 2013). In fact, the only treatment that turned out to raise compliance was a threatening message: “If you do not respond to this letter, we will contact you personally” (Fellner et al., 2013, p. 641). However, the authors did find that social normative messages had a weak positive effect in municipalities with widespread non-compliance. In municipalities with low evasion rates, however, social normative messages actually had a weak negative effect (Fellner et al., 2013). This corresponds to a paper from the Irish Office of the Revenue Commissioners, which argues that when people wrongly believe non-compliance is more prevalent than it actually is, tax compliance can be reinforced by correcting these misconceptions (Walsh, 2012). But when non-compliance is less prevalent than people perceive it to be, descriptive norms are not effective measures to alter behavior (Walsh, 2012). People’s prior beliefs might therefore play an important role in the effectiveness of descriptive norms. In addition to people’s prior beliefs other factors might play a role. For instance, altruistic preferences could affect someone’s willingness to pay for public TV or taxes in general when he or she learns that compliance is already very high (or low). This makes it very difficult to predict the effectiveness of descriptive norms upfront.

The same was shown in the previously mentioned field experiments by Ayres et al. (2012) that provided people with feedback about neighbors’ energy use. After seeing a comparison, rebound effects occurred because efficient users actually started to increase their energy use. Another study showed that providing people with peer information on retirement savings decisions actually decreased the savings of non-participants (Beshears, Choi, Laibson, Madrian, & Milkman, 2015). In fact, savings decreased when observed peer savings were higher. According to Beshears et al. (2015) people might be discouraged by such upward social comparisons. The results of a study that examined the effects of social norms on theft of wood showed that presenting individuals with descriptive norms would most likely increase theft rather than reduce it (Cialdini, et al., 2006).

The phenomenon of backfiring social norms has been captured in a well-known study on that investigated the constructive, destructive and reconstructive power of social norms (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). This paper emphasized that even though descriptive norms

may decrease undesirable behavior among individuals whose energy use was above the norm, the opposite happened among individuals who already perform the desirable behavior. In essence, the (constructive) decrease in energy consumption of some people was undermined by a (destructive) boomerang effect that led to an increase in energy use by others. Adding injunctive norms to the message proved to be ameliorating as it got rid of the boomerang effect among individuals who were already performing the desirable behavior (reconstructive). According to Schultz et al. (2007), the overall impact of a normative campaign will depend on the ratio of people who are engaged in the desirable behavior already and those who are not.

To reiterate, in various cases using descriptive norms to encourage certain behavior has yielded very promising results. In other instances, the results were less promising while in some cases using descriptive norms was even counterproductive. The possibility of backfiring descriptive norms discourages organizations to experiment with them as the results can be costly and this calls for a reliable way to predict how effective descriptive norms will be upfront.

Overall, this makes for a compelling case to test whether EUR library can use descriptive norms to increase compliance among its borrowers by means of an experiment and whether it is possible to know *so a priori* by means of a survey.

2.2 Fines

The data of this experiment, and in particular that as part of the control periods, also allowed for testing the effect of a change in EUR library's policy in January 2016. This change entailed the removal of fines that the library previously incurred for the second and third reminder, which served as penalties for borrowers who possess overdue books and who had been warned once already.

Since the dawn of time humans have looked for ways to alter undesirable behavior of others or themselves. This has fueled the discussion about the 'carrot or stick' approach, which describes ways to alter behavior by either using a reward (carrot) or penalty (stick) or a combination of the two (carrot and stick). Pigouvian subsidies and taxes for instance, are ways to deal with under-production and over-production respectively, by rewarding positive and taxing negative economic externalities (Pigou, 2013).

Past research has investigated extensively whether penalties have an effect on changing behavior. A review of the literature on deterrence and criminality concludes that "the empirical evidence is consistent with the hypothesis that punishment and other general incentives exert a deterrent effect on offenders" (Ehrlich, 1996, p. 65). Another study showed that rewards alone are relatively ineffective in increasing cooperation, that the stick treatment is slightly more effective, and that combining rewards and punishment actually has a very strong effect (Andreoni, Harbaugh, & Versterlund, 2002). This corresponds to the concept of loss aversion, which implies that "losses loom larger than corresponding

gains” (Tversky & Kahneman, 1991, p. 1039). This means that losing €100 because of a fine (stick) decreases one’s utility more than a €100 reward (carrot) increases it.

However, determining the most effective height of a penalty is very complicated. There have been prior studies on the optimal tradeoff between the probability and severity of punishment (Polinsky & Shavell, 1979; Polinsky & Shavell, 1990; Garoupa, 2001), but these provide theoretical frameworks for situations where the probability of detection does not equal 1. In this study however, that probability is 1 as the EUR library’s IT system automatically recognized when a book is due and whether it is handed in or not.

Although it is often assumed that higher fines are more deterrent than low ones, they often do not result in higher rule compliance. For instance, one study on driving offences Moffatt & Poynton (2007) did not find a significant relationship between the fine amount and the likelihood of reoffending and the authors suggest that substantial increases in fines would not be effective in deterring recidivist offenders. Although the library setting is different, the results might be comparable since the library also contacts recidivists, i.e. borrowers who already failed to hand in the books on time at least once. Furthermore, the authors conclude that “the lack of evidence for a marginal deterrent effect of fines found in this study is consistent with much of the deterrence research on punishment severity” (Moffatt & Poynton, 2007, p. 10). Moreover, one literature review stated that: “Results revealed negative statistical associations between certainty of punishment and crime rates. However, the statistical associations between severity of punishment and crime rates are considerably weaker. (...) Current research confirms earlier correlational and quasi-experimental studies and indicates consistent and significant negative correlations between the likelihood of conviction and crime rates. The data on severity effects is less impressive.” (Von Hirsch, Bottoms, Burney, & Wikstrom, 1999, p. 1).

However, sparked by Gneezy and Rustichini (2000), the effectiveness of penalties was cast into doubt altogether. Their study at Israeli kindergartens actually revealed that significantly more parents picked up their children late after the kindergartens imposed a fine as compared to when they had not (Gneezy & Rustichini, 2000). According to the authors, people felt like they no longer had to stick to social norms (being on time) since they now paid for it instead (market norms), and they speculate that the social norm goes away for a long time when it collides with a market norm. After the fines were removed parents’ behavior did not improve as there were now no fines or social norms left (Gneezy & Rustichini, 2000).

3. Hypotheses Development

The aim of this study is to find out whether compliance among a library’s borrowers who possess overdue books can be increased by presenting them with descriptive norms. This study sheds light on whether descriptive norms can be applied in the specific setting of Erasmus University Rotterdam’s

library and draws conclusions on the generalizability of descriptive norms as such. Additionally, I investigate the relationship between the severity of punishment and borrowing compliance.

3.1 EUR library's current communication flow

The university library's loan period for books is three weeks and when students do not return the book in time they will receive a first reminder¹ (hereafter referred to as R1) on the due date. Since the library's IT system focuses on books rather than borrowers, reminders are sent out for every book that is due and borrowers therefore receive multiple reminders if they possess more than one due book. A subsequent reminder (R2) follows after the second week. If the borrowers fail to take action by either handing in the book or renewing the loan, they receive additional reminders². After receiving the third reminder (R3), the borrower's library card will be blocked until the person contacts the library and hands in the book. Renewing the loan after a third reminder is no longer possible. Due to technicalities of the IT system, the first three reminders are identical and cannot be adjusted individually. After the fourth reminder (R4), a €10 fine will be incurred for administrative costs. After not responding to the fourth reminder, eventually a €50 replacement bill will be sent manually as the fifth reminder (R5), totaling the cost (per book) to €60. These costs and consequences of not taking action are summarized in Table 1 and an example of the current reminders can be found in Appendix D.

Table 1: University Library's borrowing procedure

Message	Costs
1 st reminder (R1)	Free
2 nd reminder (R2)	Free
3 rd reminder (R3)	Free & library card blocked
4 th reminder (R4)	€10.00 administrative costs
5 th reminder (R5)	€50.00 replacement costs + €10.00 administrative costs

Every day at 6.00 a.m., the first four recall messages are sent out to borrowers who have not renewed their reservation or handed in the book since the prior reminder that was sent to them the week before. All borrowers who have taken action will not receive a new reminder. Figure 1 summarizes what the library's communication looks like.

¹ 'Reminder' and 'recall message' are used interchangeably in this paper.

² Renewing the book is still possible after the first two reminders if nobody else has reserved the book.

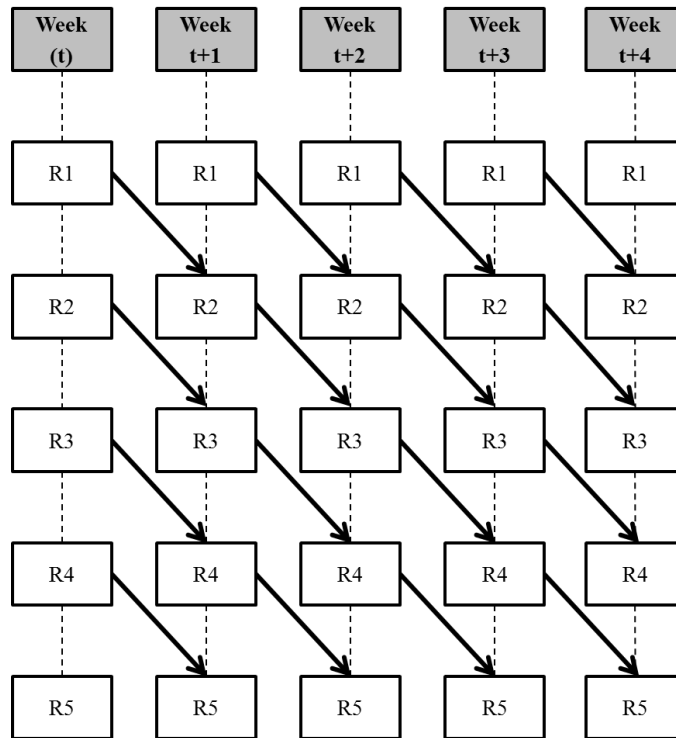


Figure 1: Weekly recall flow

Every month the library sends out thousands of reminders (R1, R2, R3 and R4) to its borrowers and urges them to renew the loan or bring back their overdue books (see: Figure 2). Unfortunately, the library does not keep track of the number of R5 messages, which are sent out manually. From Figure 2 it is clear that many borrowers ignore (multiple) reminders, either on purpose or by accident. This leads to frustration among the library's borrowers, since these books are unavailable to others and putting a book on hold (referred to as 'reserving a book' by EUR's library) is useless when someone else does not return the book. Furthermore, administrative costs are incurred because (lifting) library card blockages and sending out fifth reminders manually requires time from library staff.

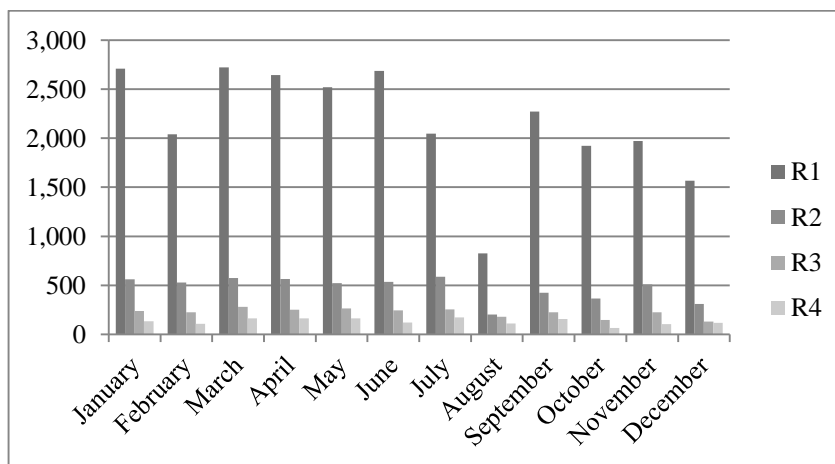


Figure 2: Monthly number of 1st, 2nd, 3rd and 4th reminders in 2015

3.2 EUR library's communication flow in 2015

There are several differences between the borrowing procedures in 2015 and those in 2016. First of all, the library eliminated the first two fines that were in place in August 2015, and replaced the one for R4 with a higher fine. Until August 2015 the first reminder was free, but the second, third and fourth reminders incurred costs of €1.00, €1.75 and €2.50, respectively. In January 2016, while the previous fines had been eliminated, a €10.00 fine was introduced for the fourth reminder. In order to find the true effect of descriptive norms it is crucial that the experiment controls for the changes in the library's borrowing procedures (see Appendix H for an overview). At the same time, this allows for testing the effectiveness eliminating the first two penalties the library has imposed on borrowers who were too late with handing in their books. This will be discussed in the fourth section (Experiment).

One other procedural change, which was also implemented at the end of 2015, was the removal of the one-week courtesy period between the due date and the first reminder for regular borrowers and the four-week courtesy period for university staff. While this may be expected to affect the number of first reminders that were sent out, it is unlikely that this affects compliance because two types of borrowers have opposite effects on compliance as a result of this change in policy. The first group consists of borrowers who have forgotten about due date but who would have remembered and taken action the next week, regardless of receiving a reminder. When the courtesy week is removed this group increases the proportion that take action after receiving a reminder. The second group consists of borrowers who want to hold on to their book for one more week after due date without taking action, regardless of receiving a reminder or not. When the courtesy week is removed, this group does not take action after receiving the first reminder and decreases compliance, as compared to when the courtesy week would still be there (because they would not have received a reminder yet). Since it is expected these two groups offset each other, it is unlikely this explains differences in compliance.

Copies of the first three reminders in 2015 can be found in Appendix A, Appendix B and Appendix C.

3.3 The book compliance proportion

Borrowing behavior can be measured both in the number of books that are due, as well as the number of people that actually hold due books. Since EUR library's current IT system keeps track of outstanding loans based on books rather than borrowers all calculations and tests are based on the number of overdue books.

As was shown in Figure 2, second recall messages are sent out per book if the owner did not take action since the first reminder, i.e. they have not renewed the book or handed it back in, third recall messages are sent to people who did not take action since the second reminder and so on. Therefore, it can be calculated how persuasive a certain reminder (e.g. R2) is by calculating the proportion of books

for which action was taken within one week afterward the reminder to the number of reminders that were sent out, which is shown step by step below.

Book compliance proportion

The total amount of first recall messages sent out in a specific week t is represented by $R1(t)$:

$$R1(t) = \text{Number of books that are overdue at week } t$$

The total amount of books that have not been renewed or handed in after receiving a first reminder in the prior week is represented by $R2(t+1)$. This equals the total amount of second reminders sent out in week $t+1$:

$$R2(t+1) = \text{Number of } R1 \text{ books that are still due at week } t+1$$

Therefore, the amount of books that have actually been handed in or have been renewed can be calculated by subtracting the amount of second recall messages ($R2$) at week $t+1$ from the total of first reminders ($R1$) at week t :

$$\text{Number of } R1 \text{ books that are not due any longer} = R1(t) - R2(t+1)$$

The persuasiveness of recall message $R1$ at time t then equals the proportion of the respective second reminders to first reminders:

$$R1(t) \text{ book compliance proportion} = \frac{R1(t) - R2(t+1)}{R1(t)}$$

For instance, if the library sends out 150 $R2$ recall messages in week 14 and it has sent out 500 $R1$ messages in week 13, the book compliance proportion would be:

$$R1(t=13) \text{ book compliance proportion} = \frac{R1(13) - R2(14)}{R1(13)} = \frac{500 - 150}{500} = \frac{7}{10} = 0.7$$

This specifies the proportion of books for which a next recall message is unnecessary (7/10) since they have been handed in or renewed. The remaining proportion exists of people whom were apparently not persuaded enough by the first reminder (3/10) to hand in their book, and who subsequently received a second reminder. That reminder's persuasiveness is thus based on the amount of borrowers took action within a week after receiving it.

Book compliance proportions can thus be calculated weekly for each reminder, except for $R4$ and $R5$ since no records have been kept about the weekly amounts of fifth reminders. Therefore, $R4$ and $R5$ are also not depicted in Figure 3.

Finally, book compliance proportions can also be calculated for each reminder over longer periods of time. The longest period that is tested for in the experiment comprises four weeks and the book compliance proportion for a reminder (e.g. R1) over a period of four weeks (e.g. week 20 up to and including week 23) looks as follows:

$$R1(20, 21, 22, 23) \text{ book compliance proportion} \\ = \frac{[R1(20) + R1(21) + R1(22) + R1(23)] - [R2(21) + R2(22) + R2(23) + R2(24)]}{[R1(20) + R1(21) + R1(22) + R1(23)]}$$

3.4 Hypotheses

As part of the EUR library's 2016 borrowing procedures, the first two fines of €1.00 and €1.75 respectively were essentially removed, while the third one was replaced by a higher one (R4 incurred €10 instead of €2.50). This means that in 2016 there were no penalties in place anymore for the first three reminders except for the membership blockage after R3. It is hypothesized that eliminating these fines deters less borrowers and hence compliance proportions will be lower. The first hypothesis is therefore as follows:

H1: Compared to incurring fines, eliminating fines for recall messages (R2 and R3) yields lower book compliance proportions (R1, R2 and R3).

Based on the existing literature, as discussed in the second section (Literature Review), it is hypothesized that recall messages are more persuasive when they include descriptive norms. This translates into higher book compliance proportions for these recall messages, as compared to the standard ones, ceteris paribus. Therefore, the second hypothesis is as follows:

H2: Adding descriptive norms to recall messages (R1, R2 and R3) yields higher book compliance proportions (R1, R2, R3) as compared to the standard recall messages.

4. Experiment

In order to find the true effects descriptive norms and the adjusted fine structure, a natural field experiment was conducted at Erasmus University Rotterdam's Library. Similar to lab experiments, field experiments can investigate causality, but unlike lab experiments they do so in a natural setting. While the subjects were unaware of being studied, this natural field experiment tested for the effects of two different treatments: 1) inclusion of descriptive norms in the library's communications to borrowers (referred to as 'norm treatment or norm period'), and 2) the effect of removing two fines as the EUR library changed its borrowing procedures (referred to as 'penalty treatment or penalty period').

A typical downside of field experiments is that there is less control over confounding variables that might affect the search for causality as compared to conventional lab experiments. This was also the

case in this field experiment. As EUR's library's current IT system did not allow for splitting up the population of borrowers it was only possible to test the effect of a single treatment message per period since interaction effects could occur if different messages were sent out at the same time. Therefore, the control treatment (week 16-23, 2015), penalty treatment (week 16-19, 2016) and norm treatment (week 20-23, 2016) were conducted in different periods and the entire borrower population was exposed to each treatment.

It is conceivable that since the three treatment periods are at different times of the year, calendar effects from public holidays, exam weeks and/or thesis deadlines play a role. For instance, the penalty treatment period's first week (week 16) was an exam week for the Erasmus School of Economics, which is EUR's largest faculty in terms of students, while there was no exam week during the norm treatment. On the other hand, the norm treatment period was closer to general thesis deadlines, for instance for the Rotterdam School of Management, which is EUR's second largest faculty.

However, statistical analysis of week 16-19 and week 20-23 in 2015 by means of a Pearson's Chi-Squared test and a two-sided Fisher's exact test revealed that although there is some weekly variation, both periods are relatively similar and that the weeks offset each other. During week 16-19 three weekly compliance proportions were significantly higher (at least at a 5% significance level) in both tests and three were significantly lower (one at a 10% and two at a 1% significance level) as compared to week 20-23. One proportion was significantly higher during week 16-19 (exactly at a 10% significance level) in the Chi-Squared test, but not in Fisher's exact test. There were no significant differences in between the two periods for the five remaining weekly compliance proportions. Furthermore, when the two periods were compared as wholes the differences in compliance proportions also offset each other³. The R1 compliance proportion was 1.12% higher in week 20-23 as compared to week 16-19, but this result was not significant. And even though the R2 compliance proportion was 15.3% higher (this was statistically significant at a 5% significance level) during week 20-23 as compared to week 16-19, the R3 compliance proportion was 25.3% lower during week 20-23 as compared to week 16-19 (this was statistically significant at a 1% significance level).

These results demonstrated that a fair comparison between the two periods was possible in 2015, and because in 2016 the exam weeks of two major faculties (Rotterdam School of Management and Erasmus School of Economics) were in the same week numbers as in 2015 and public holidays such as King's Day, Liberation Day, Ascension Thursday and Pentecost were included both in 2015 and 2016, it is assumed this also holds for 2016. Furthermore, the control period therefore functioned as a legitimate control treatment to compare the penalty treatment and the norm treatment to. In the next

³ See appendix F for more information.

section (4.1), the penalty treatment is discussed and Section 4.2 elaborates on the norm treatment and the survey that was conducted to find the most suitable treatment for the experiment.

4.1 Penalty treatment

Firstly, the penalty treatment resulted from the EUR library changing its policy, and was therefore not a treatment I instigated. On the one hand, it served as a way to control for the changes in penalties in order to find the true effect of the norm treatment. At the same time, this enabled testing for the effect of eliminating the fines for the second and third reminders of €1.00 and €1.75 respectively. The penalty treatment therefore also serves as a way to test for the first hypothesis.

The penalty treatment period (week 16-19) in 2016 was compared to the control period in 2015 (week 16-23). If the first hypothesis holds true, eliminating the fines was less effective than incurring them and compliance would be lower for R1, R2 and R3 during the penalty treatment period as compared to the control period in 2015.

4.2 Norm treatment

Secondly, to test for the second hypothesis the effect of descriptive norms the norm treatment was conducted during the norm treatment period (week 20-23) in 2016. The norm treatment period was compared to the control period in 2015 to see if the difference was larger than that between the penalty treatment period and the control period. In order to find the true effect of descriptive norms, I also compared the norm treatment period directly to the penalty treatment period, when procedures were identical.

Since it is difficult to predict the effectiveness of descriptive norms upfront, a survey served as a proxy for selecting the best descriptive norms treatment, which will be explained in the next sub-section.

4.2.1 Survey on descriptive norms

Since it was only possible to send out one treatment message containing descriptive norms, it was imperative that the treatment message was as persuasive as possible. Furthermore, there was a possibility that descriptive norms would backfire in this experiment. This problem is inherent to usage of descriptive norms in general as it is often not clear *a priori* whether it will turn out to be effective or not, or even counterproductive.

In order to solve both issues at once, an attempt to predict the effectiveness of descriptive norms upfront by means of a survey was made. This online survey served as a proxy for finding the treatment message with the greatest potential, while also predicting whether it would be successful (or counterproductive) at all. Three different messages were constructed, which have all been proven to be successful in prior studies, and adjusted in order to fit the library setting. For each of the three reminders, subjects were asked to state on a 7-point Likert scale how much they agreed with the following statement:

“This sentence convinces me to hand in the book within one week.” (1=Entirely Disagree and 7=Entirely Agree). The compliance proportions and percentages that served as descriptive norms were based on an analysis of the university library’s outstanding books on the third of May 2016.

The first message was based on a study by Hallsworth et al. (2014): “*Nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet.*”. It reads as follows:

1. 'More than eight out of ten of the University Library's books are handed in on time. You are currently in the very small minority of people who have not done so yet.'

The second message was based on research by Goldstein et al. (2008): “*The majority of guests reuse their towels.*”. It reads as follows:

2. 'The majority of books are handed in on time by our borrowers.'

The third and final option was based on a study by Kettle et al. (2014): “*According to our records, 64.5% of Guatemalans declared their income tax for the year 2013 on time. You are part of the minority of Guatemalans who are yet to declare for this tax.*”. It reads as follows:

3. 'According to our records, 86.5% of the University Library's books are handed in on time. You are part of the minority of the University Library's borrowers who are yet to do so.'

The online survey was conducted among 105 respondents, who studied at a university or had done so in the past and/or who owned a university library card. This target group is relatively similar to that of the EUR library, which lends out its books to a varied population of students, university staff and alumni, as well as other members. I specifically excluded individuals who were still studying at Erasmus University Rotterdam from taking the survey, so potential future borrowers were not exposed to the treatment message prior to the treatment period of the experiment.

Since all subjects expressed their opinion about all three messages the experiment uses a within-subject design. In order to find which message outperforms the others in persuasiveness, the data was analyzed using a Wilcoxon signed-rank test in Stata. This is a (non-parametric) statistical method to test for consistent differences between pairs of observations. Since these three messages have not been compared before, it cannot be hypothesized that one will outperform the other and hence a two-sided test was used to analyze the data.

Table 2: Survey statistics of the two-sided Wilcoxon signed-rank test

Message	Mean	Message	Mean	P-values
1.	4.133	2.	3.448	0.000***
2.	3.448	3.	4.343	0.000***
3.	4.343	1.	4.133	0.1311

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

The Wilcoxon signed-rank test showed that there were significant differences between the first and the second message as well as between the second and the third message. On average, the first message scored significantly higher as compared to the second message in persuading people to hand in their books within one week. This result is statistically significant at a 1% significance level ($p = 0.000$), ceteris paribus. On average, also the third message scored significantly higher as compared to the second message, and likewise this result is statistically significant at a 1% significance level ($p = 0.000$), ceteris paribus. There was no significant difference ($p = 0.131$) between the first and third message but the third message's mean (4.343) was higher and its standard deviation (1.580) was lower as compared to the first message's respective mean (4.133) and standard deviation (1.776).

Nevertheless, only one message could be applied in the field experiment. It was assumed that 4, when respondents neither agree nor disagree with the statement that the message was effective, was the neutral point of the survey. It was therefore imperative that the treatment message would score higher than 4 for it to be effective. The Wilcoxon signed-rank test showed that the third message was the only message which scored significantly higher ($p=0.022$) than 4, and it was therefore selected as the treatment message in the field experiment⁴.

4.2.2 Recall messages with descriptive norms

Based on the statistical analysis of the survey in the previous paragraph, the following message was selected as the norm treatment message: *'According to our records, 86.5% of the University Library's books are handed in on time. You are part of the minority of the University Library's borrowers who are yet to do so.'* The treatment in this experiment entailed altering the library's communication with its borrowers who possess overdue books by adding descriptive norms. By emphasizing that the large majority of borrowers, including students and university staff as well as external borrowers do return their books on time, it was tested if borrowers who own due books are persuaded that they should stick to the norm as well. If the second hypothesis holds true this would result in a faster return of books by borrowers who were already late, which would be signaled by lower amounts of consecutive reminders.

During the treatment period, the treatment message was incorporated in all of the existing reminders that the library sends out, including those for books that are due one week up to those who are due four

⁴ See Appendix G for more details.

weeks (R1, R2, R3 and R4). An example of the treatment reminders, which include the descriptive norms, can be found in Appendix E. Apart from adding the descriptive norms at the beginning of the reminder, the recall messages during the treatment period (2016) were identical to those in control period II in 2016. The treatment reminders were also similar to those in 2015, except for the content regarding the fines and the descriptive norms. It is crucial to note that isolating which aspect of a message produces a particular effect is difficult. Prior studies have shown that factors such as the tone, vocabulary and length of the message can all play a role (Jackson, 1992). To control for these factors as much as possible the treatment message that was added to the reminders, i.e. the two phrases that addressed descriptive norms, were kept very short. As mentioned previously, the design and the other content of the treatment message were not changed.

5. Results of Study

Ultimately the compliance proportions during the three different treatment periods, namely the control period in 2015 and the penalty treatment and norm treatment periods in 2016 were compared to find the effect of including descriptive norms in recall messages and the alteration of the library’s fines. The compliance proportions over each respective period are summarized in Figure 4.

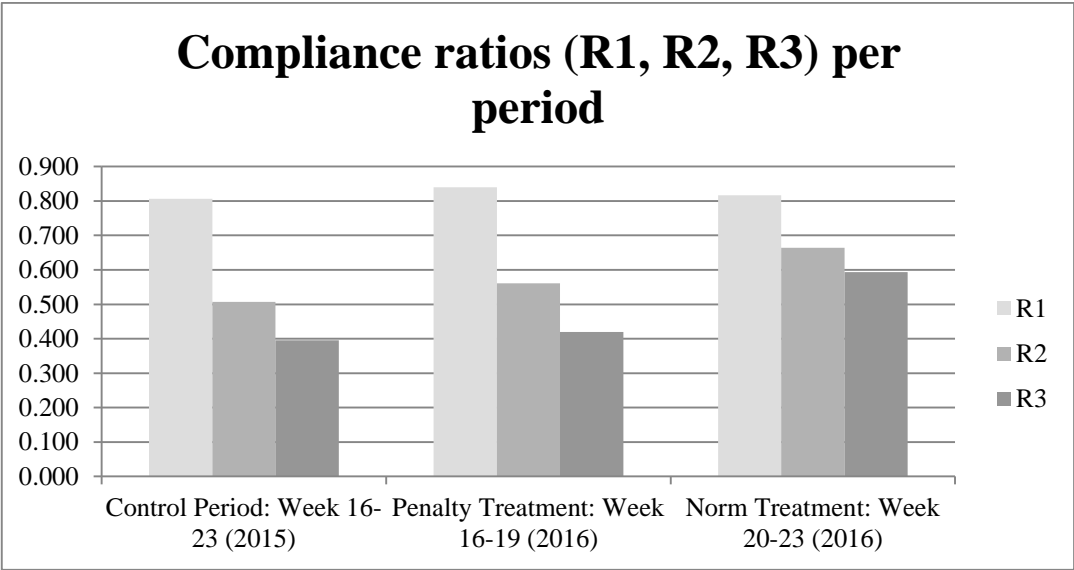


Figure 4: Compliance proportions per period

Identifying any significant differences between the three different periods requires statistical analysis of their respective compliance proportions. Deviations of the compliance proportion were theoretically possible in either direction, both for the penalty treatment and the norm treatment and therefore two-tailed tests were used. On the one hand, Pearson’s Chi-squared was suitable but its significance value is only an approximation, even though that is mostly a problem in small sample sizes (Lancaster & Eugene, 2005). Fisher’s exact test, on the other hand, is more exact but some have argued

that it is too conservative, which means one gets a significant result, $P < 0.10$ for example, less than 10% of the time (McDonald, 2009). Therefore, both the Pearson's Chi-squared test and a two-tailed Fisher's exact test were used to determine whether there were significant differences between the book compliance proportions per period.

In section 5.1, I made a comparison between the penalty treatment period in 2016 and the control period in 2015 in order to find the effect of the altered fines. In section 5.2, the norm treatment period was compared to the control period as well, in order to find the effectiveness of descriptive norms. Additionally, it includes a third comparison between the treatment period and control period in 2016 to find the exact difference between the two.

5.1 Penalty treatment

In order to find the effect of the adjusted fines on compliance the penalty treatment period was compared to the control period in 2015. As can be seen in Table 3, all three compliance proportions were higher during the penalty treatment period in 2016 than during the control period in 2015, but not all results were significant. There were for instance, no significant differences between the R3 compliance proportions during the penalty treatment period and the control period. But the R1 compliance proportion was significantly higher during the penalty treatment period as compared to the control period and this was statistically significant at a 1% significance level. Likewise, the R2 compliance proportion was significantly higher during the penalty treatment period as compared to the control period, and this was statistically significant at a 10% significance level.

In fact, during the penalty treatment the R1 and R2 compliance proportions were respectively 4.9% and 10.7% higher as compared to the control treatment. In contrast to the first hypothesis, which stated that "Compared to incurring fines, eliminating fines for recall messages (R2 and R3) yields lower book compliance proportions (R1, R2 and R3)." these results suggest that eliminating fines yielded higher compliance proportions for R1 and R2 and that R3 did not change significantly.

Table 3: P-values Pearson's Chi-squared test and Fisher's exact test

Control Period (2015)	Penalty Treatment Period (2016)	P-values Pearson's Chi-Squared test	P-values Fisher's exact test
R1 (3891/4823) = 0.807	R1 (2664/3173) = 0.840	0.000***	0.000***
R2 (495/905) = 0.507	R2 (264/471) = 0.561	0.060*	0.061*
R3 (182/460) = 0.396	R3 (84/200) = 0.420	0.558	0.605

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

5.2 Norm treatment

In a similar fashion, the norm treatment period was compared to the control period in 2015. Table 4 shows that once more all compliance ratios were higher during the treatment period than the control

period, and that again not all of these results were significant. The R2 compliance proportion during the norm treatment period was significantly higher as compared to the control period and this was statistically significant at a 1% significance level. The R3 compliance proportion was also significantly higher at a 1% significance level during the penalty treatment period as compared to the control period. Compared to the control treatment, during the norm treatment the R2 compliance proportion was 31.2% higher and the R3 compliance proportion was 49.7% higher. The R1 compliance proportion was 1.23% higher during the norm treatment period than during the control period, but this result was not statistically significant.

Table 4: P-values Pearson's Chi-squared test and Fisher's exact test

Control Period (2015)	Norm Treatment Period (2016)	P-values Pearson's Chi-Squared test	P-values Fisher's exact test
R1 (3891/4823) = 0.807	R1 (2638/3231) = 0.817	0.276	0.283
R2 (495/905) = 0.507	R2 (410/617) = 0.665	0.000***	0.000***
R3 (182/460) = 0.396	R3 (143/241) = 0.593	0.000***	0.000***

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

5.3 Norm treatment vs penalty treatment

At a first glance, the results of the penalty treatment and the norm treatment appear to be relatively similar because both feature two compliance ratios that were significantly higher during the treatment period as compared to the control period. But the increases in percentages of two compliance proportions were much higher during the norm treatment than during the penalty treatment.

Therefore, a direct comparison of the penalty treatment period (week 16-19) and the norm treatment period (week 20-23) was required to determine the exact difference in effectiveness. As can be seen in Table 5, the R1 compliance proportion was higher in the penalty treatment period (0.840) as compared to the norm treatment period (0.817) and this result is statistically significant at a 1% significance level. But both the R2 and R3 compliance proportions were higher in the norm treatment period (0.665 and 0.593 respectively) as compared to the penalty treatment period (0.561 and 0.420 respectively). Both of these results are significant at a 1% significance level.

Overall, there were significant differences between the norm treatment and the penalty treatment for all three compliance proportions. As compared to the penalty treatment, during the norm treatment the R1 proportion was 2.73% lower, while the R2 proportion was 18.5% higher. The biggest difference was the R3 proportion, which was 41.2% higher during the norm treatment than during the penalty treatment.

Table 5: P-values Pearson's Chi-squared test and Fisher's exact test

Penalty Treatment Period (2016)	Norm Treatment Period (2016)	P-values Pearson's Chi-Squared test	P-values Fisher's exact test
R1 (2664/3173) = 0.840	R1 (2638/3231) = 0.817	0.008***	0.008***
R2 (264/471) = 0.561	R2 (410/617) = 0.665	0.000***	0.000***
R3 (84/200) = 0.420	R3 (143/241) = 0.593	0.000***	0.000***

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

In order to examine the differences, I made a more detailed comparison and drew four weekly pairwise comparisons between the compliance proportions in the penalty treatment period and those in the norm treatment period for each of the three reminders. The compliance proportions in each of the four weeks of the penalty treatment period were compared to their counterparts of the norm treatment period. Thus, for each of the three reminders the following comparisons were made: week 16 and week 20; week 17 and week 21; week 18 and week 22; and week 19 and 23. The twelve compliance proportions are summarized in Table 6.

As Table 6 shows, six out of the twelve compliance proportions were significantly higher (at least at a 5% significance level) during the norm treatment period as compared to the penalty treatment period. The two R1 compliance proportions in week 16 and 17 (during the penalty treatment period) were significantly higher as compared to week 20 and week 21 respectively (during the norm treatment period). These results are significant at a 1%- and 5%-significance level respectively. The other four compliance proportions were not significantly different in the treatment period as compared to the control period.

To conclude, considering the weeks individually does not make a difference in terms of conclusions, as it shows once more that the majority of weekly compliance proportions were higher during the norm treatment period than during the penalty treatment period. Therefore, the results are in line with my second hypothesis, namely "Adding descriptive norms to recall messages (R1, R2 and R3) yields higher book compliance proportions (R1, R2, R3) as compared to the standard recall messages."

Table 6: P-values Pearson's Chi-squared test and Fisher's exact test

Week	Penalty Treatment Period (2016)	Norm Treatment Period (2016)	P-values Pearson's Chi-Squared test	P-values Fisher's exact test
16 / 20	R1(632/728) = 0.868	R1 (634/865) = 0.733	0.000***	0.000***
	R2 (62/104) = 0.596	R2 (105/142) = 0.739	0.017**	0.019**
	R3(27/76) = 0.355	R3 (45/83) = 0.542	0.018**	0.025**
17 / 21	R1 (717/828) = 0.866	R1 (636/774) = 0.822	0.015**	0.016**
	R2 (70/96) = 0.729	R2 (158/231) = 0.684	0.481	0.509
	R3(25/42) = 0.595	R3 (20/37) = 0.541	0.624	0.655
18 / 22	R1 (687/847) = 0.811	R1 (677/783) = 0.865	0.003***	0.004***
	R2 (55/111) = 0.496	R2 (90/138) = 0.652	0.013**	0.014**
	R3 (12/26) = 0.462	R3 (44/73) = 0.603	0.212	0.253
19 / 23	R1 (628/770) = 0.816	R1 (691/809) = 0.854	0.039**	0.042**
	R2 (77/160) = 0.481	R2 (57/106) = 0.534	0.367	0.383
	R3 (20/56) = 0.357	R3 (34/48) = 0.708	0.000***	0.000***

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

6. Discussion

6.1 Implications

One could see important differences between both of the two treatment periods and the control period, as well as between the treatment periods themselves. Both the penalty treatment and the norm treatment increased the compliance proportions, which means that more borrowers followed up on these treatment messages by extending the overdue book or handing it in. It appears that the library's communications with its borrowers is most persuasive when the first two fines were eliminated and descriptive norms were included in its reminders. The results carry three major implications for the applications of penalties and descriptive norms.

First of all, the results from this study indicate that removing penalties that have been in place for years can be an effective measure to increase compliance, and it appears that a fine indeed comes at a price. However, this conclusion must be drawn with caution. It is also possible that the higher fine for R4 deterred borrowers so much that this increased compliance among borrowers who were one or two weeks due. This would then contradict a substantial amount of literature that found no evidence for an effect of penalty severity on compliance. What is particularly interesting, however, is that one could then expect the impact of such a fine to be highest among borrowers who already received R3 and who are about to actually receive this fine if they do not take action. Yet, the R3 compliance proportion was not significantly higher during the penalty treatment period as compared to the 2015 control period. This

means that either the higher fine did not contribute to the increase in compliance, or that it deters borrowers early in the borrowing process, but not right before the threat of the high fine is close. This topic therefore deserves more attention in future research, which is discussed in Section 6.3 (Recommendations for Future Research).

The second implication is that descriptive norms appear to be effective means to further increase compliance when penalties are already in place. Even though the R1 proportion was actually 2.73% lower during the norm treatment than during the penalty treatment, the R2 and R3 proportions increased tremendously, in comparison to both the norm treatment (31.2 and 49.7% respectively) and the penalty treatment (18.5% and 41.2% respectively). Since people only receive the first reminder after they are already late with handing in the book, the R1 group might include a lot of people who really did want to hand it in but simply forgot. Perhaps the first reminder should therefore be seen as some kind of warning, which is not given before due date, and social norms may therefore not be as effective. The results suggest that people only feel like they violate the norm when they received a second or a third reminder, and not yet after the first one. Arguably, one could then expect the effect of descriptive norms to be largest in the second and the third reminder, when borrowers have not taken action and have consciously violated the norm at least once.

As a result, descriptive norms turn out to be a cost-effective way of improving the persuasiveness of a library's communications to its borrowers and increasing compliance. Yet, several authors (for instance: Bohner & Schlüter, 2014; Fellner, Sausgruber, & Traxler, 2013) have questioned the generalizability of descriptive norms. This study however, presents yet another example where descriptive norms are effective. This strengthens the case that descriptive norms are (at least oftentimes) indeed effective measures to alter behavior, as the Behavioural Insights Teams so heavily promotes in its EAST-model (The Behavioural Insights Team, 2014) and other literature (such as Ayres et al., 2012; Berkowitz, 2004; Köbis, 2015; Walsh, 2012) suggests. Since the costs of this intervention were extremely low, it is worthwhile to look into further applications elsewhere, both at the university and in policy measures elsewhere, and these are addressed in Section 6.3 (Recommendations for Future Research).

The effectiveness of descriptive norms has theoretical implications about the concepts of social norms and market norms as well. In his book *Predictably Irrational*, Dan Ariely (2008) builds on the idea that we live in two worlds, one where social norms dominate and one where market norms dominate. Ariely (2008) points out complications that arise when these are mixed. Furthermore, the aforementioned study, Gneezy and Rustichini (2000) reported that newly introduced fines (market norms) collided with existing social norms and speculated that when this happens social norms would stay away for a long time. Even though this study showed that fines might indeed set a price on undesirable behavior, it also found that addressing descriptive norms is an easy way to reestablish social

norms, if they actually vanished, by simply presenting them to the library's borrowers. The new fine system may have enhanced this process because it focuses more on social norms than the old system, since it only imposes a fine after multiple appeals to individuals' values did not work.

The final implication of this study, is that it provides a compelling case in favor of using a survey in combination with a field experiment to predict and test for the effectiveness of descriptive norms. As was discussed in the second section (Literature Review), the difficulty of using social norms is that it unclear clear upfront whether a norm will be effective, ineffective or even counterproductive. This study applied a very cheap way to estimate the effect of descriptive norms *a priori* by means of an online survey. Although it remains unclear whether the selected treatment was indeed more effective than the other two options, it turned out to significantly increase compliance nonetheless. It seems therefore that a survey is a helpful tool since it correctly predicted that these descriptive norms would increase compliance (and not backfire). Eliminating or at least reducing the chances of counterproductive results by using such surveys could prompt the usage of surveys as proxies for experiments with descriptive norms. This in turn, could prove to be a solution for the uncertainty about the effect *a priori* and encourage both individuals and organizations to experiment more with applications of descriptive norms.

6.2 Limitations

In addition to emphasizing the favorable results it is crucial to note the limitations of this experiment. First of all, as was mentioned previously, this field experiment was conducted in the natural setting of the EUR library. Since the library's IT system did not allow sending out different messages simultaneously a between-subject design was not possible. Instead, both the treatment and the control messages were sent out to the entire population at different points in time. Therefore, the *ceteris paribus* condition was very difficult to maintain completely and several confounding factors could have played a role. It is therefore recommended that a follow-up experiment is conducted if the library's new IT system allows for splitting up the borrower population, so all of the following limitations can be controlled for.

First of all, the University Library's borrowing procedures changed between 2015 and 2016⁵. The courtesy period for instance, was removed and since the recall messages contained information regarding the borrowing procedures the reminders in 2015 were very similar but not completely identical to those in 2016. In the comparison between the control period and the penalty treatment period this might have interacted with the effect of the fines that were removed. In the comparison between the control period and the norm treatment period it might have interacted with the effect of descriptive norms. Nevertheless, this interaction effect is at least partially controlled for because of the comparison

⁵ See appendix H for a summary of changes in the library's borrowing procedures.

between the penalty treatment period and the norm treatment period because the procedures, fines and messages were completely identical.

Secondly, severely bad weather in one of the two periods or other unobservables might have discouraged (more) people from going to campus, where EUR library is located. However, as all control and treatment periods consisted of at least four weeks it is unlikely borrowers could not find a chance to hand in their books. Moreover, the treatment period was compared to two periods in two different years and the results were similar, namely that descriptive norms enhanced compliance among the library's borrowers.

6.3 Recommendations for Future Research

In the light of the limitations above, I touch upon some relevant topics for future research. First of all, more research could be done on the optimal severity for penalties such as fines in order to increase rule compliance. Various prior studies have argued that penalty severity does not affect compliance. Although this study showed that compliance right before the fine did not increase, one single, higher fine might have increased compliance in an early stage already. It also might be that for any single fine to be effective a certain threshold (such as €10) must be met, or that people only perceive fines as fair and follow up on them after multiple reminders. Since determining the optimal severity of penalties is crucial for policymakers additional research on this topic would be valuable.

Secondly, this study found that descriptive norms can (further) increase compliance when penalties are already in place. Replication in other (public) libraries, rental companies or non-borrowing related environments may be useful in generalizing the findings of this study on rule compliance. It is recommended that in addition to descriptive (social) norms also injunctive (social) norms and personal norms are researched in order to find optimal combinations of (social) norms in combination with existing penalties.

Testing for the effects of penalties and descriptive norms in *a priori* reminders is a third relevant topic for future research. Participants in this study had already failed to hand in their books on time, either deliberately or by mistake. It might very well be that these individuals, in particular the latter group, would be affected by *a priori* reminders since Apesteguia et al. (2013) showed that these increase compliance. It would then be interesting to see what the effect of mentioning penalties, descriptive norms, or a combination of the two, in those *a priori* reminders would be. As reminders upfront are sent out to everyone, including people who would have handed in their books on time anyway, caution should be exercised to potential rebound effects and destructive boomerang effects (as described by Ayres et al., 2012; Schultz, 2007). Again, a survey might serve as a reliable predictor of success.

7. Conclusions

This study has taken an alternative perspective on the carrot-or-stick discussion by studying the effects of removing existing fines and presenting non-compliant borrowers with descriptive norms.

It turns out that removing multiple, low fines for handing in books one or two weeks yielded higher compliance proportions among those borrowers. Although it is possible a higher fine for being three weeks contributed to this, this same fine did not deter the borrowers who were about to receive it. This has important implications for any policymaker who aims to deter potential non-compliant individuals in different stages of being due, ranging from public libraries to rental companies.

Furthermore, including descriptive norms in communications is a cost-effective way to further increase compliance when a penalty is already in place. Market norms do not necessarily collide with (descriptive) social norms and in fact a combination of the two can be the best way to maximize compliance. Presenting non-compliant borrowers with descriptive norms is an effective way to reinvigorate these social norms. The library setting in this study adds another example where descriptive norms were applied successfully and helps to build a case for using and testing for applications of descriptive norms elsewhere.

Moreover, the effectiveness of the norm indicates that using an online survey helps to eliminate at least a part of the risk and costs of ineffective and backfiring descriptive norms. This benefits both individuals and organizations that were unwilling to conduct such experiments thus far because descriptive norms could backfire, and this in turn facilitates future research on this topic.

8. Bibliography

- Andreoni, J., Harbaugh, W., & Versterlund, L. (2002, March 14). *The Carrot or the Stick: Rewards, Punishments and Cooperation*. University of Oregon.
- Apestequia, J., Funk, P., & Iriberry, N. (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review*, 64, 266-284.
- Ariely, D. (2008). *Predictably irrational*. New York: HarperCollins.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership, and men*, 5, 222-236.
- Ayres, I., Raseman, S., & Shih, A. (2012). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *Journal of Law, Economics, and Organization*, ews020.
- Berkowitz, A. D. (2004). *The social norms approach: Theory, research, and annotated bibliography*.
- Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., & Milkman, K. (2015). The effect of providing peer information on retirement savings decisions. *The Journal of finance*, 70(3), 1161-1201.
- Blumenthal, M., Christian, C., Slemrod, J., & Smith, M. G. (2001). *Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota*. National Tax Association.
- Bohner, G., & Schlüter, L. E. (2014). A room with a viewpoint revisited: Descriptive norms and hotel guests' towel reuse behavior. *PloS one*, 9(8).
- Cadena, X., & Schoar, A. (2011). Remembering to pay? Reminders vs. financial incentives for loan payments. *National Bureau of Economic Research*, No.w17020.
- Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, 115, 144-160.
- Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social influence*, 1(1), 3-15.
- Cialdini, R., Reno, R., & Kallgren, C. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6), 1015.

- Dolan, P., Hallsworth, M., Halpern, D., & King, D. (2010). *MindSpace: Influencing behaviour through public policy*. London, UK: Cabinet Office.
- Ehrlich, I. (1996). Crime, punishment, and the market for offenses. *The Journal of Economic Perspectives*, 10(1), 43-67.
- Fellner, G., Sausgruber, R., & Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 634-660.
- Garoupa, N. (2001). Optimal magnitude and probability of fines. *European Economic Review*, 45(9), 1765-1771.
- Gneezy, U., & Rustichini, A. (2000). Fine is a price, a. *J. Legal Stud.* 29, 1.
- Goldsmith, R., & Clark, R. (2012). Materialism, status consumption, and consumer independence. *The Journal of social psychology*, 152(1), 43-60.
- Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, 35(3), 472-482.
- Haines, M. P., Barker, G. P., & Rice, R. (2003). Using social norms to reduce alcohol and tobacco use in two midwestern high schools. In H. W. Perkins, *The social norms approach to preventing school and college age substance abuse: A handbook for educators, counselors, and clinicians* (pp. 235-244). Jossey-Bass.
- Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2014). *The behavioralist as tax collector: using natural field experiments to enhance tax compliance*. Cambridge, US: NATIONAL BUREAU OF ECONOMIC RESEARCH.
- Haynes, L., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: developing public policy with randomised controlled trials. *Cabinet Office - Behavioural Insights Team*.
- Haynes, L., Greeb, D., Gallagher, R., John, P., & Torgerson, D. (2013). Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *Journal of Policy Analysis and Management*, 32(4), 718-730.
- Jackson, S. (1992). *Message effects research: Principles of design and analysis*. New York: Guilford Press.
- Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*.

- Kettle, S., Hernandez, M., Ruda, S., & Sanders, M. (2014). *Behavioural Interventions in Tax Compliance: Evidence from Guatemala*. US: Internal Revenue Service.
- Köbis, N. C., van Prooijen, J. W., Righetti, F., & Van Lange, P. A. (2015). “Who Doesn’t?”—The Impact of Descriptive Norms on Corruption. *PloS one*, *10*(6), e0131830.
- Lancaster, H. O., & Eugene, S. (2005). *Chi-Square Distribution*. John Wiley & Sons, Ltd.
- Leibenstein, H. (1950). Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. *The Quarterly Journal of Economics*, *64* (2), 183-207.
- Long, D. S., Fox, N. R., & York, T. (2007). *Calculated futures: theology, ethics, and economics*. Baylor University Press.
- McDonald, J. H. (2009). *Handbook of biological statistics (Vol. 2, pp. 77-85)*. Baltimore: MD: Sparky House Publishing.
- Mehrabian, A. (1998). Effects of poll reports on voter preferences. *Journal of Applied Social Psychology* *28*(23), 2119–2130.
- Moffatt, S., & Poynton, S. (2007). Deterrent Effect of Higher Fines on Recidivism: Driving Offences, The. *BOSCAR NSW Crime and Justice Bulletins*, 15.
- Nolan, J., Cialdini, P., Goldstein, N., & Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin* *34*, 913-923.
- Pigou, A. (2013). *The economics of welfare*. Palgrave Macmillan.
- Polinsky, A. M., & Shavell, S. (1990). A note on optimal fines when wealth varies among individuals (No. w3232). *National Bureau of Economic Research*.
- Polinsky, A., & Shavell, S. (1979). The optimal tradeoff between the probability and magnitude of fines. *The American Economic Review*, *69*(5), 880-891.
- Reese, G., Loew, K., & Steffgen, G. (2014). A towel less: Social norms enhance pro-environmental behavior in hotels. *The Journal of social psychology*, *154*(2), 97-100.
- Rothschild, D., & Malhotra, N. (2014). Are public opinion polls self-fulfilling prophecies? *Research & Politics*, *1*(2).
- Samson, A. (2015). *The Behavioral Economics Guide 2015 (with an introduction by Dan Ariely)*. Retrieved from www.behavioraleconomics.com

- Schultz, P., Nolan, J., Cialdini, R., Goldstein, N., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429-434.
- Speake, J. (2015). *Oxford dictionary of proverbs*. Oxford: OUP Oxford.
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, *4*(3), 199-214.
- Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral decision making*, *12*(3), 183.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
- The Behavioural Insights Team. (2014). *EAST: Four simple ways to apply behavioural insights*.
- The Behavioural Insights Team. (2014, February 4). *The Behavioural Insights Team – a social purpose company*. Retrieved from <http://www.behaviouralinsights.co.uk/news/the-behavioural-insights-team-a-social-purpose-company/>
- The Behavioural Insights Team. (2015). *Update Report 2013-2015*. The Behavioural Insights Team.
- Torgler, B. (2004). Moral suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Economics of Governance*.
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, 1039-1061.
- Von Hirsch, A., Bottoms, A. E., Burney, E., & Wikstrom, P. O. (1999). *Criminal deterrence and sentence severity: An analysis of recent research*. Oxford: Hart.
- Walsh, K. (2012). Understanding Taxpayer Behaviour - New Opportunities for Tax Administration. *The Economic and Social Review*, Vol. 43, pp. 451-475.
- Weatherburn, D., & Moffatt, S. (2011). The specific deterrent effect of higher fines on drink-driving offenders. *British journal of criminology*, *51*(5), 789-803.
- Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, *26*(6), 862-883.
- Wilde, O., & Bristow, J. (2006). *The picture of dorian gray*. Oxford: OUP Oxford.

9. Appendices

Appendix A: The first reminder (R1) during control period I and II (2015)

Uitleenbureau-bibliotheek@rug.nl

Erasmus University Library

21-05-2014

Dear ~~Mr/Mrs~~ UB Groningen
POSTBUS 559

9700 AN GRONINGEN

Recall notice: 1** recall (0,00)

This is a recall for an overdue book. To avoid additional costs you may renew the loan by using the renew function in the University Library Catalogue:

<http://lbs-prd.ubib.eur.nl:8080/LBS WEB/borrower/borrower.htm?USR=888&BES=1&LAN=NE>

If the book has been reserved, the loan cannot be renewed.

Hoe test ik de snelheid van een rappel

16-04-2014

298 G 89

0.00

If you have any questions regarding this mail, please contact our Virtual Desk: http://www.eur.nl/ub/english/forms/ask_your_question/

Appendix B: The second reminder (R2) during the control period I and II (2015)

Uitleenbureau-bibliotheek@rug.nl

Erasmus University Library

21-05-2014

Dear ~~Mr/Mrs~~ UB Groningen
POSTBUS 559

9700 AN GRONINGEN

Recall notice: 2nd recall(1,00)

This is a recall for an overdue book. To avoid additional costs you may renew the loan by using the renew function in the University Library Catalogue:

<http://lbs-prd.ubib.eur.nl:8080/LBS WEB/borrower/borrower.htm?USR=888&BES=1&LAN=NE>

If the book has been reserved, the loan cannot be renewed.

Hoe test ik de snelheid van een rappel
16-04-2014
298 G 89
1.00

If you have any questions regarding this mail, please contact our Virtual Desk: http://www.eur.nl/ub/english/forms/ask_your_question/

Appendix C: The third reminder (R3) during control period I and II (2015)

Uitleenbureau-bibliotheek@rug.nl

Erasmus University Library

21-05-2014

Dear ~~Mr/Mrs~~ UB Groningen
POSTBUS 559

9700 AN GRONINGEN

Recall notice: 3rd recall(1,75)

This is a recall for an overdue book. To avoid additional costs you may renew the loan by using the renew function in the University Library Catalogue:

<http://lbs-prd.ubib.eur.nl:8080/LBS WEB/borrower/borrower.htm?USR=888&BES=1&LAN=NE>

If the book has been reserved, the loan cannot be renewed.

Hoe test ik de snelheid van een rappel

16-04-2014

298 G 89

1,75

If you have any questions regarding this mail, please contact our Virtual Desk: http://www.eur.nl/ub/english/forms/ask_your_question/

Appendix D: Example of a reminder (R1, R2 & R3) during control period III (2016)

kers@ubib.eur.nl

500001219

Erasmus University Library

31-03-2016

Dear Mr/Mrs J.C. KERS
W V B 12

3e rappel (niet verlengbaar) / 3rd recall (not renewable)

This is a recall for an overdue book. To avoid costs you may renew the loan by using the renew function in the University Library Catalogue:

http://lbs-eur.oclc.org:8080/LBS_WEB/borrower/borrower.htm?USR=888&BES=1&LAN=EN

If the book has been reserved, the loan cannot be renewed.

Signature : 999 A 99
Title : Hoe test ik de snelheid van een rappel
Author : Uitlening
Date of expiration : 16-03-2016
Recall costs (EUR) : 0.00

Recall Procedure:

If you do not return the book in time you will receive a recall after the expiration date and subsequent recalls after the second, third and fourth week.

Fourteen days after the fourth recall you will receive a bill for the replacement of the book. The costs per book are:

1st recall: free of charge

2nd recall: free of charge

3rd recall: free of charge and library card blocked 4th recall: EUR 10.00 Replacement bill: EUR 60.00
(replacement costs EUR 50.00 + recall costs EUR 10.00)

If you have any questions regarding this mail, please contact our Virtual Desk:

<http://service.ubib.eur.nl/questionpoint/askaquestion.html>

University Library - Erasmus University - Burgemeester Oudlaan 50 - 3062 PA Rotterdam

Appendix E: Example of a reminder (R1, R2 and R3) during the treatment period (2016)

kers@ubib.eur.nl
500001219

Erasmus University Library

14-05-2016

Dear Mr/Mrs J.C. KERS
W V B 12

'According to our records, 86.5% of the University Library's books are handed in on time. You are part of the minority of the University Library's borrowers who are yet to do so.'

1e rappel (niet verlengbaar) / 1st recall (not renewable)

This is a recall for an overdue book. To avoid costs you may renew the loan by using the renew function in the University Library Catalogue:
http://lbs-eur.oclc.org:8080/LBS_WEB/borrower/borrower.htm?USR=888&BES=1&LAN=EN

If the book has been reserved, the loan cannot be renewed.

Signature : 999 A 99
Title : Hoe test ik de snelheid van een rappel
Author : Uitlening
Date of expiration : 13-05-2016
Recall costs (EUR) : 0.00

Recall Procedure:

If you do not return the book in time you will receive a recall after the expiration date and subsequent recalls after the second, third and fourth week. Fourteen days after the fourth recall you will receive a bill for the replacement of the book. The costs per book are:

1st recall: free of charge
2nd recall: free of charge
3rd recall: free of charge and library card blocked
4th recall: EUR 10.00
Replacement bill: EUR 60.00 (replacement costs EUR 50.00 + recall costs EUR 10.00)

If you have any questions regarding this mail, please contact our Virtual Desk:
<http://service.ubib.eur.nl/questionpoint/askaquestion.html>

University Library - Erasmus University - Burgemeester Oudlaan 50 - 3062 PA Rotterdam

Appendix F: Control period 2015

Table 7: P-values

Week 16-19 2015	Week 20-23 2015	P-values Pearson's Chi- Squared test	P-values Pearson's Fisher's exact test
R1 (1910/2381) = 0.802	R1 (1981/2442) = 0.811	0.427	0.444
R2 (206/438) = 0.470	R2 (253/467) = 0.542	0.032**	0.033**
R3 (96/209) = 0.459	R3 (86/251) = 0.343	0.011**	0.013**

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

Table 8: P-values

Week	Week 16-19 2015	Week 20-23 2015	P-values Pearson's Chi- Squared test	P-values Pearson's Fisher's exact test
16/20	R1 (435/533) = 0.787	R1(471/570) = 0.826	0.092*	0.097*
	R2 (39/80) = 0.488	R2 (61/113) = 0.540	0.474	0.474
	R3 (38/69) = 0.551	R3(29/92) = 0.315	0.003***	0.004***
17/21	R1 (586/664) = 0.883	R1 (427/550) = 0.7764	0.000***	0.000***
	R2 (58/118) = 0.492	R2 (48/99) = 0.485	0.922	1.000
	R3 (21/41) = 0.512	R3(15/52) = 0.289	0.028**	0.034**
18/22	R1 (495/657) = 0.753	R1 (605/737) = 0.821	0.002***	0.002***
	R2 (39/78) = 0.500	R2 (67/12) = 0.545	0.536	0.564
	R3 (20/60) = 0.333	R3 (22/51) = 0.431	0.289	0.329
19/23	R1 (394/507) = 0.777	R1 (478/585) = 0.817	0.1*	0.112
	R2 (70/162) = 0.432	R2 (77/132) = 0.583	0.010***	0.014**
	R3 (17/39) = 0.436	R3 (20/56) = 0.357	0.439	0.523

*=Significant at an $\alpha=0.1$ level, **=Significant at an $\alpha=0.05$ level and ***=Significant at an $\alpha=0.01$ level.

Appendix G: Wilcoxon signed-rank test

Message	Sample size	Mean	Standard Dev.	Min.	Max.	Neutral point	Prob > z
'More than eight out of ten of the University Library's books are handed in on time. You are currently in the very small minority of people who have not done so yet.'	105	4.133	1.776	1	7	4	0.4187
'The majority of books are handed in on time by our borrowers.'	105	3.448	1.569	1	7	4	0.0009
'According to our records, 86.5% of the University Library's books are handed in on time. You are part of the minority of the University Library's borrowers who are yet to do so.'	105	4.343	1.580	1	7	4	0.0217

Appendix H: Overview of changes in the University Library's borrowing procedure

Reminder	2015 Procedure (since 3 June 2014)	Update August 2015	2016 Procedure (since 11 January)
First reminder	Free	Free	Free & removal of the 1-week courtesy period (and the 4-week one for staff)
Second reminder	€1.00	Free	Free
Third reminder	€1.75 + card blocked	Free	Free + card blocked
Fourth reminder	€2.50	Free	€10
Fifth reminder	€50 replacement costs + €12.50 administration costs + €5.25 reminder costs	-	€50