The London School of Economics and Political Science

Optimising Eco-feedback Design

Ukasha Ramli

A thesis submitted to the Department of Psychological and Behavioural Science of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London May 2021

DECLARATION

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 31,068 words.

STATEMENT OF CO-AUTHORED WORK

I confirm that Paper 3 was co-authored with Kate Laffan. I confirm that I conducted 100% of the design and implementation of the research work, 100% of the data analysis, and 50% of the writing.

STATEMENT OF INCLUSION OF PREVIOUS WORK

I can confirm that Paper 1 has been published in the journal Sustainability 13(5) in March 2021.
https://doi.org/10.3390/su13052796

ABSTRACT

Eco-feedback has become one of the most popular behavioural interventions for promoting household water and energy conservation. Since its inception, it has been adopted by various companies and governments around the world as one path to addressing climate change. Due to its ubiquity, eco-feedback interventions have been designed in various ways, potentially leading to heterogeneity in its treatment effects. This thesis investigates the different components of eco-feedback interventions, and how these can moderate its treatment effects

Through four field experiments, I study the moderating effects of duration, frequency, medium, and to an extent, content of eco-feedback interventions. I find that 1) eco-feedback is effective at reducing household water/energy consumption across various contexts, achieving between 1-2% reduction in consumption, 2) the effects of the treatment attenuates over time once the treatment has ceased, 3) the medium by which the feedback is delivered is critical to its effectiveness, 4) delivering feedback for both water and energy at the same time may have a negative effects, and 5) the treatment effects are heterogeneous, mostly based on a household's baseline consumption.

Insights from this thesis should help inform the design of future eco-feedback interventions to better maximise its effects in the most cost effective manner.

TABLE OF CONTENTS

TABLES, FIGURES AND EQUATIONS

ACKNOWLEDGEMENTS

For supporting me academically:

Abhijeet Singh

Lee Crawfurd

Rachael Meager

Paul Dolan

Kate Laffan

For supporting in implementing my studies:

Julien Lancha

Patrice Guillouzic

Pauline Marcou

Doug Jones

Stefan Irava

1. INTRODUCTION

Feedback is the process of providing information to people about their behaviour to help reinforce or change those behaviours (Skinner 1938; Bandura 1969). It is beneficial to goal pursuit as it facilitates the selection and adoption of goals (Fishbach & Finkelstein, 2012). Feedback also plays an important role in regulating behaviour (Carver & Scheier, 1982; Carver & Scheier, 2012) and has subsequently been used as an effective form of intervention for behaviour change across various fields including education (Schunk and Swartz 1993; Hanna 1976), management (Dahling et al., 2015), environment, better known as eco-feedback (Brook, 2011; Abrahamse et al., 2007), addiction (Cunningham et al., 2012) and public health (Fuller et al. 2012). Out of these, eco-feedback has probably received the most attention in both academia and policy (Karlin et al. 2015; Nisa et al. 2019; Nauges & Whittington, 2019). Despite its widespread application though, its design and relatedly, its effectiveness has varied widely. The aim of the current PhD is therefore to better understand the dynamics of effective feedback interventions, specifically in the way it is applied to household water and energy consumption. The PhD will be broken down into three separate papers that will each try and build upon the literature of feedback interventions by identifying how different designs can moderate their efficacy. Paper 1 looks at the long term effects of eco-feedback interventions after delivery of the treatment has ceased, paper 2 looks at the medium and frequency at which the treatment is delivered, and paper 3 looks at ways to deliver feedback for both water and energy at the same time. Empirical evidence for the PhD will derive from field-based experiments in the UK and a country in the Middle East. The research here draws on studies and insights from across various fields including psychology, economics, human-computer design, data science and environmental research. The insights gained here can be used to prevent wasteful practices of delivering these interventions, while maximising their effectiveness.

1.1. Why feedback interventions

There are three main reasons why feedback interventions have become an important approach to behaviour change in recent times. Firstly, the current ubiquity of sensors and tracking devices means that there is a proliferation of behavioural data that is available to be utilised for creating more accurate feedback than ever before. In the health sector, health monitoring wristbands and apps such as Fitbit and Jawbone have resulted in increases in motivation and behaviour through the feedback provided by these devices (Asimakopoulos et al., 2016). In the US, 69% of its population are tracking at least one health behaviour, with 14% using one of these specialised health monitors (Fox & Duggan, 2012). Likewise, the proliferation of smart meters for household water and energy consumption have created new opportunities for households to monitor their consumption behaviours with more granularity than ever before (Karlin et al., 2015). In education, new technologies allow for greater tracking and feedback of learning progress amongst school children (Muralidharan et al., 2016). Digital devices are tracking more of our behaviour than ever before, and the available data gathered have made it significantly easier for society to gain more knowledge regarding human behaviours than ever before. This torrent of digital behavioural data has created lots of new opportunities to psychologically engage with people (Harari et al., 2016).

Even in the absence of technical sensors and monitors, people are motivated to track their behaviours by voluntarily recording their actions, thoughts and feelings into digital devices as part of a growing movement known as the Quantified Self (Lupton, 2016). The Quantified Self is a movement whereby individuals engage in the tracking of their own biological, physical, behavioural, or environmental information, with the main objective of self-improvement (Li & Guo, 2016). By tracking themselves, individuals can then experiment with changes to their lifestyle to help improve it in some way. This highlights people's motivation to engage with information about themselves and is a second reason as to why feedback interventions are of significance. The attraction of self-knowledge means that people are likely to engage more with feedback interventions that contain self-knowledge than other non-personalised behaviour change messaging. This can be widely seen in the research on feedback-seeking behaviour (e.g., Crommelinck & Anseel, 2013; VandeWalle, 2003). This level of personalisation of information available also highlights another benefit of feedback interventions. Research has shown that behaviour change interventions that can reflect a degree of personalisation are more effective than those that are seemingly generic (Service et al. 2015).

A final reason for the importance of feedback interventions is that there is a wide body of evidence that not only highlights its effectiveness (e.g., Harkin et al. 2016; Karlin et al. 2015; Hermsen et al., 2016, Fischer, 2008) but especially its superiority over other forms of behaviour change interventions. For example, a review of 122 behaviour change interventions for healthy eating found that interventions that utilised techniques related to self-regulation (e.g., self-monitoring and feedback) were significantly more effective than other interventions such as rewards or encouragement (Michie et al., 2009). Furthermore, the Education Endowment Foundation (2013) highlighted feedback interventions as one of the most impactful interventions for improving education. One reason why other forms of interventions may not have been as successful is because other popular models such as the Theory of Planned Behaviour (Ajzen, 1991), Theory of Reasoned Action (Fishbein & Ajzen, 1975) and the Model of Interpersonal Behaviour (Triandis, 1977), focus on the role of goal intention, which alone is not enough to ensure goal attainment and behaviour change (Gollwitzer & Sheeran, 2006; Sheeran, 2002; Sheeran & Webb, 2011). An argument here is that these models do not explain the process that transforms intention formation into behaviour change (de Bruin et. al, 2012), whereas, as will be demonstrated here, the goal pursuit driven by feedback does lead to behaviour change.

Looking at pro-environmental behavioural interventions, feedback interventions, especially those utilising social normative comparisons, have been found to be highly effective. In a meta-analysis of pro-environmental behavioural interventions, Nisa et al (2019) found this form of intervention to be the second most effective and reliable form of intervention after choice architect interventions. However, as described by the authors, there has so far been a limited number of areas where changing choice architectures have been implemented, with most studies focusing on switching to non-meat products. This is likely because 1) not all behaviours can be changed by intervening on its choice architecture, and 2) some behaviours are too personalised and context specific to practically implement a choice architecture intervention that is widely applicable. When it comes to household water and energy consumption, choice architecture interventions often focus on a single consumption behaviour, for example, reducing the use of energy by setting lower/higher temperatures as the default on thermostats for heating/cooling homes (Brown & Johnstone, 2013). It would not, however, be able to address all other related energy consuming behaviours, without implementing additional interventions on every single point where behaviour interacts with consumption. On the other hand, feedback interventions work by generating the motivation for behaviour change on a higher or more abstract level of consumption, which allows individuals to tackle each behaviour that is unique to their environment themselves. In more practical terms, feedback interventions can easily leverage the consumption data that is generated from household water and energy meters to deliver an intervention to a vast number of people quickly and consistently. The simplicity in implementation is what allows feedback interventions to be easily implemented on a vast number of people continuously and may also be a reason for its popularity. Studies that do utilise choice architecture tend to have much smaller sample sizes, possibly due to the challenges in implementing them widely. Furthermore, since feedback interventions are so broadly applied, the heterogeneity of the treatment effects can help to explain the smaller effect sizes, which will be explored in each paper. Finally, feedback interventions for household energy and water consumption are also very cost effective as demonstrated in paper 2 of this thesis but has also been identified in many others (e.g., Allcott, 2011, Nauges & Whittington, 2019).

1.2. What are feedback interventions

Early research into the use of feedback to change behaviour date back to the study of knowledge results (KR) (e.g., Jones, 1910; Judd, 1905), where participants were provided with the results of their experimental task, which generally positively correlated with performance. Subsequently, in early behaviourism research, KR was used as the reinforcement or punishment in operant conditioning (Skinner, 1938; Thorndike, 1927). Bandura (1969) then expanded on the concept of feedback by not only looking at the results of a behaviour, but also the process of engaging that behaviour (e.g., "You went to the gym twice this week"), referencing the behaviour to a goal (e.g., "You are 5kg away from reaching your ideal weight") and referencing results to comparators (e.g., "You've lost the most weight compared to your peers"). Feedback interventions can, therefore, be formally defined as, actions taken by an external agent to provide information regarding some aspect of one's task performance or behaviour with the goal of changing or improving that performance or behaviour (Kluger & DeNisi 1996).

Since its early iterations, feedback interventions have developed into far more sophisticated forms of interventions. As will be discussed below, depending on the technology and data available, feedback interventions can be delivered through a multitude of formats and channels, as well as with different framing and context. Therefore, the design of feedback interventions can be broken down into two main components. The first component is how the feedback is delivered. This includes the format or medium in which the feedback is delivered, the frequency at which it is delivered and the duration of the overall intervention. The second component is the content of the feedback itself. This includes the measurement of the behaviour, the granularity of that measure and how the feedback is framed (e.g., compared against a benchmark).

1.3. Two main functions

In addition to the two components of feedback intervention design, there are also two main functions of feedback interventions, learning and motivation (Ammons 1956). Feedback for learning typically involves providing detailed task level information related to one's performance that is to be used to improve performance and/or learning in some way. For example, basic feedback might state that "you can run 100m in 13 seconds", whereas feedback for performance would add "your strides are too long and can be made shorter for more cadence". These types of feedback can be applied to improving performance in sporting activities (Garcia-Gonzalez et al., 2013), pseudoword reading aloud tasks (Mattheiss et al., 2017), surgical training (Zahiri et al., 2017), tutoring of robots (Vollmer et. al, 2014) and education (Hattie & Timperley, 2017). On the other hand, feedback for motivation typically involves providing information regarding a behaviour framed in a way that promotes motivation through various processes (Kluger & DeNisi, 1996). For example, providing feedback on one's gambling behaviour alone may not be effective, but framing it with feedback of normative gambling behaviours can motivate corrective behaviour change (Cunningham et al., 2012).

While the two functions are distinct in their objective and approach, they are often intertwined. For example, changes in motivation have been found to facilitate the processing of performance/learning type feedback (DePasque & Tricomi, 2015). Furthermore, feedback for learning can increase knowledge and awareness of one's own behaviours and its impact, and that education alone can sometimes be sufficient for behaviour change (e.g., Alberts et al., 2011). Therefore, while the focus here is on the effect of feedback on motivation, the role of feedback on learning will also be a relevant component.

1.4. Self-regulation

The basic mechanism of how feedback leads to behaviour change as proposed by various behaviour change theory is rooted in the concept of self-regulation (Fishbach & Finkelstein, 2012; Orehek et al., 2011). When presented with feedback, individuals compare their current behaviour or performance against some set goal or standard. When a discrepancy or error signal between that standard and feedback is detected, individuals become motivated to reduce the discrepancy (Bandura & Cervone, 1986; Carver & Scheier, 2012; Matsui et al., 1981; Champion & Lord, 1982). Progress is then monitored through feedback and behaviour is changed until the feedback of current behaviour is in line with the set standard. This unit is known as a feedback loop and has its roots in the transdisciplinary field of cybernetics (Carver & Scheier, 2012). Feedback loops are therefore made up of four elements; an input function, a reference standard, a comparison process, and an output (MacKay, 1996; Miller et al., 1960; Powers, 1973; Carver & Scheier, 2012). In the context of feedback interventions, the input function is the feedback received, the reference standard is the set goal, and the output is the subsequent change in behaviour. This theory of self-regulation of motivation and goal-directed behaviour is an important component of many theories in psychology including, Social Cognitive Theory (Bandura, 1986), Control theory (Carver & Scheier, 1982), Goal Setting theory (Locke & Latham, 1990), Perceptual Control theory (Powers, 1973), Goal Systems theory (Kruglanski et. al 2002), Feedback Intervention theory (Kluger & DeNisi, 1996) and Embodied Cognition (Balcetis & Shana, 2009).

1.5. Goal commitment and motivation

Goals can serve as reference points for behaviour (Heath et al., 1999), and as such play an important role in feedback interventions. These goals may include being as good or better than oneself, others, or some idealised standard. Without a goal, there is nothing for feedback to be compared against. Two factors determine goal commitment, attractiveness and expectancy of goal attainment (Locke & Latham, 2002; Liberman & Förster, 2008; Klein, et al., 1999). Goal attainment can be made important or attractive through its social value (Cialdini & Trost, 1998), through incentives (Latham 2001), through the setting of the goal by an authoritative figure (Durham et al., 1997) or by self-participation (Dossett et al., 1979). In regard to the expectancy of goal attainment or self-efficacy, if a goal seems unattainable due to its difficulty or complexity, it is less likely to be adopted and pursued (Bandura, 1977). Self-efficacy can also inform the subjective utility of achieving the goal (Atkinson, 1964).

Goal commitment, however, is not a necessary precursor to feedback interventions. Goals can also be defined and adopted through feedback itself. Attitudes research finds that people have a desire to be consistent and to express stable preferences over time (Festinger, 1957; Cialdini et al., 1995; Bem, 1972; Harmon-Jones & Mills, 1999). Additionally, self-perception theory argues that people learn about themselves and their preferences by observing their own behaviour (Bem, 1972). Therefore, when positive feedback regarding a seemingly unattended goal is presented to an individual, that individual may rationalise that the goal was, in fact, valuable to the self. For example, someone who finds out that they recycle more than the average person might infer that they value recycling or the environment, more so than if they had received negative feedback regarding their recycling behaviour. Relatedly, positive feedback helps individuals assess their level of commitment towards that goal as they internalise or integrate the new goal to their self-concept (Ryan & Deci, 2000; Fishbach et al., 2006).

Feedback can also signal the value of a goal by influencing an individual's sense of self-efficacy. As suggested by Social Cognitive Theory, perceived self-efficacy motivates behaviour by creating the view that a goal is something achievable but needs to be mastered (Bandura, 1977). For example, employees who received positive feedback regarding their work performance developed a strong sense of self-efficacy and subsequently demonstrated greater motivation on their work-related goals (Audia et al., 2000). Beyond goal commitment, feedback can also drive the revision of goals as it relates to self-efficacy. Participants receiving positive feedback on anagrams exercises were more likely to self-report increases in self-efficacy and subsequently increased their goal levels (Tolli & Schmidt, 2008). Feedback from an external agent regarding a previously unattended goal could also signal the importance or attractiveness of that goal, therefore leading to goal commitment. This would be in line with previous findings regarding the messenger effects (Dolan et. al 2010; Chaiken 1980; Pornpitakpan 2004). Finally, feedback regarding other's behaviours can also signal a social norm, which can then lead the individual to adopt the related goal (Cialdini & Trost, 1998).

1.6. Feedback seeking behaviour and monitoring goal progress

Monitoring goal progress is an essential process within the feedback loop. Whether an individual monitors their goal progress when presented with feedback can determine the effectiveness of the feedback intervention. Therefore, the facilitation of feedback seeking behaviour plays an important role in the success of feedback interventions as it encourages self-motivated engagement with the feedback. The literature makes a distinction between active and passive monitoring where the latter is unstructured and is not deliberate (e.g., noticing clothes no longer fit when trying to lose weight), while the former involves actively seeking out feedback (e.g., checking weight on a scale) (Harkin, et al., 2016). Four self-motives have been identified to explain the drive for active feedback seeking behaviour (Anseel et al., 2007; Crommelinck & Anseel, 2013). These motives are similar to those found for motivating the search for self-knowledge and self-evaluation (Sedikides & Strube, 1997). 1) Self-assessment is the desire to accurately know the truth about oneself to reduce uncertainty or for some moral obligations (Brown, 1991). 2) Self-improvement is the desire to use feedback to learn and improve performance on a selected goal. 3) Self-enhancement is the desire to seek out feedback that provides favourable self-views, thus boosting self-esteem (Swann, 1990). 4) Self-verification is the desire to ensure our self-view is aligned with how others perceive us (Swann, 1990), as well as to confirm that our self-view is an accurate reflection of ourselves as in the case of cognitive dissonance (Festinger, 1957).

Each of these motives can determine how much an individual would choose to engage with feedback. For example, an individual with a strong self-enhancement motive that is expecting negative feedback may avoid feedback to preserve their self-image. This active avoidance of feedback is known as the 'Ostrich Problem' and has been suggested to occur due to a conflict between different self-motives (Webb et al., 2013). For example, an individual with the goal of saving money may end up avoiding checking their bank balance because doing so would risk challenging their enhanced self-view. Alternatively, someone with the self-perception of being pro-environmental may avoid feedback regarding their energy consumption to avoid dissonance. Feedback seeking behaviour and its related motives are therefore an important consideration when designing feedback interventions. Due to the multitude of combinations any self-motives any individual can hold it is however very difficult to take these factors into consideration.

1.7. Variability of feedback interventions

Despite its long history, results from feedback intervention studies and its impact on behaviour change has been mixed (Ilgen et al., 1979; Salmoni et al. 1984; Balcazar et al., 1985; Karlin et al. 2015; Harkin et al. 2016; Hermsen et al. 2016; Fischer 2008). For example, a meta-analysis of 607 effect sizes by Kluger and DeNisi (1996) found that two-thirds of interventions studied were effective at changing behaviour, while a third had an opposite effect. By comparison, a more recent review of feedback interventions using digital technologies found that 59 out of the 72 studies reviewed showed a positive effect of disrupting habitual behaviours (Hermsen et al. 2016). These differences in the effect of feedback on behaviour change have been attributed to the varying elements that determine the content of the feedback interventions as well as the way the feedback is delivered. The content of feedback interventions refers to any elements related to how the feedback is framed. This includes the measurement of the behaviour, the granularity of the data and most importantly the goal or reference used to compare the feedback against. The delivery elements refer to the operational components required for delivering the content of the feedback to individuals. This includes the frequency of the feedback, the medium in which it is delivered, and the duration of the intervention. Several meta-analyses and systematic reviews have attempted to document the difference in the effects of these elements with varying results (Fischer, 2008; Harkin et. al, 2016; Karlin et al. 2015; Hermsen et al. 2016; Kluger & DeNisi, 1996).

One of the aims of the current PhD is, therefore, to test these different components within the same context to determine which component of feedback intervention is most effective at encouraging behaviour change. Furthermore, a common issue amongst the various reviews and meta-analysis of feedback interventions is the lack of studies that utilised controlled experimental methods and clear measures. For example, out of the 519 studies initially identified for a meta-analysis of feedback on quality of care, only 19 had met the criteria (Hysong 2009). Therefore, an additional aim of the PhD is to provide additional evidence of feedback interventions that utilise rigorous experimental designs.

A final potential cause of variability of feedback effectiveness may also be down to differences in the audience or recipients of the feedback. For example, a study of feedback interventions for reducing energy found that the recipients who were more politically liberal were more engaged and responsive to the feedback messaging then than those who were more politically conservative. To address this potential source of variability, a tailored approach that matches differences in feedback components with differences in audience characteristic is proposed.

2. Feedback Content

2.1. Standards

The feedback literature distinguishes between three types of standards: 1) self-referential feedback, which is typically based on the personal history of the behaviour in question (e.g. "you used more water than last week"), 2) norms based feedback, which provides a comparison to a social norm (e.g. "you used more water than your neighbours") and 3) goal based feedback which provides a comparison against some evaluative measure or desired target (e.g. "you used less water than the ideal goal of X litres").

2.1.1. Self-referential feedback

Self-referential feedback, also known as ipsative feedback, uses the initial state, behaviour or performance as the reference point within a feedback loop. In this sense, any progress achieved today is considered as positive feedback but doesn't truly have an end state because progress could continue forever. Therefore, it can be argued that self-referential feedback does not contain a specific goal. This is likely why descriptive feedback has often been found to be the least effective standard used in feedback interventions (e.g., Karlin et al. 2015; Fischer, 2008). The absence of a specific goal can lead to ambiguity and too much variance in performance as individuals have no reference point to direct their behaviour towards (Locke et al., 1989; Wright & Kachmar, 1994).

Furthermore, goal specificity influences how goal progress affects motivation (Wallace & Etkin, 2018). According to the goal gradient hypothesis (Kivetz et al., 2006) and the theory of goals as reference points (Heath, et al. 1999), as individuals move closer to their reference point (i.e., end goal), marginal progress seem more impactful and increases motivation. In the absence of a specific goal, however, it was found that individuals default to using their initial state as a reference point. Therefore, they begin where the reference point is close to goal progress, and therefore motivation is high, but as progress increases and they move away from the initial state reference point, motivation decreases (Wallace & Etkin, 2018). Descriptive feedback not only fails to provide a specific goal, but it actively uses the initial state as a reference point for feedback and goal pursuit.

Descriptive feedback has been found to be effective within learning and education interventions as it helps to track progress and promotes self-efficacy (Hughes, et al., 2014; Hughes, 2011). Within the eco-feedback literature, however, it remains unclear as to whether descriptive feedback is superior against the other two feedback standards (Hermsen et al. 2016; Karlin et al. 2015; Fischer 2008).

2.1.2. Normative feedback

According to Feedback Intervention Theory, social norms feedback would be ineffective at regulating behaviour (Kluger & DeNisi, 1996). It is suggested that any feedback that potentially threatens the self, especially if negative, is likely to be less effective than feedback that solely focuses on the task. Despite this, the use of social norms as a feedback standard has been very popular, with a large body of evidence that demonstrates its success in changing behaviour (for a review see: Nolan, Schultz et al., 2008). It has been successful in various areas including alcohol consumption (McCambridge et al. 2013), recycling (Schultz, 1999), household energy consumption (Allcott, 2011) and political engagement (Gerber & Rogers, 2009).

At the core of social norms theory are the concepts of pluralistic ignorance and false consensus, which suggest that we wrongly assume that our attitudes and behaviours are different or similar to others, respectively (Berkowitz, 2005). Normative feedback, therefore, works by overcoming pluralistic ignorance through informing the individual of what normative behaviour is, therefore setting the goal of aligning one's behaviour with the social norm. Descriptive social norms have consistently been found to be a strong motivator for behaviour, and so using social norms as a reference standard helps to facilitate goal adoption. Even when individuals rated it as being the least motivating, social norms were still found to produce more behaviour change than when other non-feedback messaging, such as pro-environmental messaging, was used (Nolan et al., 2008).

One limitation of normative feedback is that the goal being set will always depend on actual social norms. Therefore, if the social norm is not necessarily the ideal behaviour or is not a sufficient end goal for the individual, normative feedback is unable to motivate individuals any further. Relatedly, gaining positive feedback by easily aligning behaviour with social norms may lead to a 'coasting' effect whereby the individual reduces their effort and motivation towards the overall goal because they feel like they have achieved it by being in line with the social norms. This is where having better control over the setting of the goal may be advantageous.

2.1.3. Goal-setting

Goal-setting here refers to the use of any other set target as a reference standard for the feedback loop. The distinction between this goal and that of the normative feedback is that this goal typically reflects an ideal or desired goal, as opposed to what is socially prevalent. Combining goal setting and feedback has been found to be highly effective (e.g., Locke, 1991;

Michie et al. 2009; Abrahamse, et al. 2007; Fischer 2008; Darby 2006; McCalley & Midden, 2002). As previously discussed, there are various ways in and sources from which this goal can be set. Assuming the target behaviour is within a domain of a generally accepted goal area (e.g., health, financial savings, environmental), the setting of a specific standard coupled with feedback can be motivating as individuals strive to reduce feedback-standard discrepancy.

It is difficult to draw conclusions regarding the most effective reference standard to use for feedback comparison because few studies have attempted to directly compare the different reference standards. Of those that have, the results have been mixed with some finding that norms based were more effective than goal-based feedback, and vice versa (e.g., Siero et al., 1996; Burgers et al., 2015).

The literature reviewed highlights an inability to gauge a standard in which to inform behaviour. This is especially the case when no other standard is available to measure progress against, or that the messenger of the standard is seen as untrustworthy. It may well be that the norm based feedback may simply act as a reference point for behaviour and that the main motivating factor of feedback is simply, in line with Control theory, the desire to reduce the feedback-standard gap.

2.2. Measurement

Another element important to the content and framing of feedback interventions is the metric used to measure behaviour. There are various ways in which behaviours can be measured. For example, in the context of energy conservation feedback can be framed as CO_2 emissions or as the monetary value of energy consumption. The measurement used in feedback can potentially affect the importance of the goal. Again, in the context of energy conservation, an individual may value financial savings more than their impact on the environment. Therefore, they may be more motivated by measures that represent goals that are more important to them than others. Another way that measurement can moderate the effect of feedback is through the magnitude of the feedback-standard gap that it can convey. Research in numerical processing has shown that the different ways distances between pairs of numbers are presented can influence the evaluation and judgment of those values (Krajcsi & Kojouharova, 2017; Tao et al., 2017). For example, when evaluating the discrepancy between a score and an endpoint of a scale, individuals view the discrepancy as smaller when the scale is large (e.g., 0 to 100) than when it is small (e.g., from 0 to

10) (Tao et al. 2017). In the context of feedback, depending on the measurement scale used, the feedback-standard discrepancy could be perceived as larger or smaller. This is especially the case when the numerical representation of the discrepancy of two different measurements might be different in size. For example, again in the context energy conservation, a feedback-standard discrepancy presented in kilowatts per hour might be 400 kWh, whereas the financial equivalent of that amount might only be £3. These differences in the perceived magnitude of discrepancy could then have an effect on perceived self-efficacy towards the goal whereby the seemingly larger 400 kWh progress would be perceived as harder to achieve than the £3 progress. This is especially the case in abstract measures such as energy consumption where it is harder to determine how consumption directly translates to these measures. These differences in scaling could also affect perceived goal gradient. Depending on which measure is used, if the magnitude of the feedback-standard discrepancy is perceived as being different, individuals might feel they are closer to the goal and therefore increase their motivation as proposed by the goal gradient hypothesis.

One solution to resolving these issues would be to adopt the use of percentages. Nolan et al. (2008) found a change in behaviour in energy conservation when using percentages of consumption rather than a kWh measure but had not compared the two directly. Framing effects have also been found in the use of percentages. For example, people have been found to perceive a large percentage figure of a small subset (e.g., 90% of 10) as being of a greater magnitude than a small percentage figure of a large subset (e.g., 1% of 900) (Li & Chapman, 2013). Additionally, framing biases can also affect the perception of magnitude when the percentage difference is framed as a loss or a gain. For example, an end-point of 1500 can be framed as 50% more than 1000, or the 1000 can be framed as 33% less than 1500. While the difference remains the same, the perceived magnitude is found to be different (Kruger & Vargas, 2008). Therefore, even the use of percentage can influence the perceived magnitude, and subsequently the self-efficacy and goal gradient of feedback.

Another aspect of measurement is regarding the granularity at which behaviour measured. Granularity refers to the level of specificity that behaviour is measured. In the context of weight-loss, feedback granularity could either be focused on the number of calories consumed per meal or focussed on the body mass index. The increased precision gained through greater granularity has the potential of providing more learning (Sanguinetti et al., 2018; Mangiapanello, 2015). For example, by receiving high granular feedback about which appliances consume the most energy from the habitual use of it, individuals can gain significant progress in reducing energy consumption by only changing habits regarding that one appliance. Higher granularity can also be beneficial as the feedback provided can appear as more personalised and therefore more engaging. The potential risk, however, is that the high level of granularity and personalisation may frustrate individuals as they reflect on the issues of privacy. Granularity can also act as a proxy of psychological distance. In this regard, high-level construal, a more abstract or less granular measure of behaviour, is thought to be more effective at engaging self-control and self-regulation (Fujita & Carnervale, 2012). Looking at the moderating effects of granularity, Karlin et. al (2015) did not find a significant effect, whilst Fischer (2008) highlighted two studies where granularity did have a moderating effect on consumption.

The current literature does not have much to propose regarding the role of measurement. The various meta-analysis on feedback intervention again finds widely varying results on the moderating role of measurement. It is therefore of interest to better understand how the measurement used in feedback can have an effect on behaviour change. This is especially the case since measurement can have quite varying effects on the perceived magnitude of the feedback-standard discrepancy, and thus motivation.

2.3. Feedback intervention operational components

The second main component of feedback interventions that have been found to have a moderating effect is the technical operational components. These are related to the way in which the feedback intervention is delivered to individuals. There are three distinct operational components to consider, medium, frequency, and duration.

2.3.1. Medium

Feedback can be delivered through a variety of mediums from traditional analogue channels such as paper or feedback delivered by a person, to digital channels such as emails, websites, smartphone apps, wearables, in-home displays or even smart speakers. The main differences brought by the different medium include the content or information it can display and its accessibility for engagement. While the content should be assumed to be the same across mediums in any study trying to determine the effect of the medium, the accessibility can have much more immediate effects on feedback. The medium in which feedback is delivered can determine how easy or how likely feedback is to be engaged with. A wearable device is always on and instantly accessible, whereas paper-based feedback delivered to the home requires the individual to be at home and have checked their mail to receive the feedback. Even amongst digital mediums, an in-home display of energy consumption will always be present and can create saliency during the relevant events of energy consumption in the home. Whereas an email is only accessible when an individual decides to check their emails, potentially away from home and out of context. In this sense, feedback medium can have a related effect on feedback frequency whereby the temporal delay between behaviour and feedback is determined by how accessible the feedback is. Previous studies in the context of residential energy feedback suggest that feedback delivered through a digital medium are more effective than analogue, paper medium (Vassileva et al., 2012; Fischer, 2008; Karlin et al. 2015). Auditory feedback was found to be less effective than visual feedback, mostly because visual feedback could contain more details (Hoggan & Brewster, 2010).

2.3.2. Frequency

As previously mentioned, the frequency in which feedback is delivered can have an effect on subsequent behaviour change. Frequency refers to how often feedback is delivered to the individual. By extension, it can also reflect the temporal proximity or latency between an action and the subsequent feedback regarding that action. This is not always the case, however, as the updating of data for feedback could occur at a higher frequency than the intervention itself. This can be dependent on the medium in which the feedback is delivered. For example, a wearable device is generally always on and so the feedback is continuous. By comparison, a household electric meter might take recordings of consumption every hour, but feedback may only be delivered monthly via a paper report due to operational limitations. In the latter example, the feedback provided could still also highlight the hourly changes in behaviour for the individual to reflect upon.

It is generally assumed that increased frequency of feedback helps improve learning and task performance (Salmoni et al. 1984; Darby, 2006; Fischer, 2008; Opitz et al., 2011). It is thought that with increased frequency comes more information for the individual to learn from and to develop strategies to improve performance. Greater learning comes from the fact that individuals are able to more closely link an action with specific feedback in time. Greater frequency of feedback could also create greater saliency of the target behaviour. Within the context of eco-feedback, it was found that while behaviour change was found shortly after participants had received feedback, the behaviour change decayed in between each feedback report leading to a backsliding in behaviour (Allcott, 2011).

Darby (2006) and Fischer (2008) found that greater frequency led to greater energy savings, while Karlin et. al (2015) did not find this effect. As these were meta-analyses comparing experiments with different designs, the inconsistency in findings could be due to the moderating effect of the medium in which the feedback is delivered, as suggested above. Immediate feedback has also been found to be beneficial in learning contexts (Dihoff et al., 2004; Opitz et al., 2011). An alternative theory, however, suggests that increased frequency is only beneficial up to the point in which it risks overwhelming an individual's cognitive capacity, and thus reducing task effort (Lam et al., 2011). As feedback frequency increases, individuals must respond and process the feedback information and engaging self-regulating processes more frequently, which may consume more cognitive resources. Lam et al. (2011) found that feedback frequency exhibits an inverted-U relationship with performance. This relationship, however, was not suggested by the various meta-analysis. Too much feedback could also cause habituation to the motivating effects of feedback intervention or more simply could lead to disengagement with feedback as too much information leads to diminishing value of the feedback.

2.3.3. Duration

Duration refers to the length of time for which the feedback intervention is delivered. This can be under the control of an external agent (e.g., experimenter, employer, doctor, etc.) or under the control of the individual (e.g., The individual may decide to stop wearing the wearable device and disengage from its feedback). Here, the focus will be on the duration set by the external agent, but the findings would still be applicable to the individual control. In the context of eco-feedback, Fischer (2008) did not find a clear indication of this, Karlin et. al (2015) concluded that duration had a significant effect, whereby the longer a feedback intervention ran for, the more likely behaviour change occurred. The duration studied in these reviews were generally between 1 to 12 months. It was noted however that after a certain period, engagement with feedback decreased which subsequently led to an increase in energy consumption. Additionally, in a large scale eco-feedback study that sought to reduce energy consumption of five million households across the United States, Allcott (2011), found that households that were randomly selected to have their feedback interventions discontinued after two years of treatment, still persisted in their behaviour change, suggesting the formation of new consumption habits.

It should be noted, however, that the necessary duration of a feedback intervention to encourage behaviour change is also confounded by the behaviour being targeted. Not all behaviour is equally challenging to change and therefore those that are easier to change may require less time than behaviour habits that are difficult to break (Lally & Gardner, 2013; Hermsen et al. 2016). Therefore, the duration of an intervention can have an important effect on the persistence of the behaviour change.

As demonstrated here, these seemingly trivial operational components could play a role in the variability in the success of feedback interventions. These design considerations of feedback medium, frequency and duration, therefore, need to be studied further to better understand its moderating role on feedback interventions.

2.4. Individual differences and heterogeneity

Looking beyond the moderating effects of the content and operational components of feedback interventions, an additional explanation for the varying effects of feedback interventions focuses on the differences in the recipients of the interventions. As complex agents, there is a multitude of biological, environmental, or psychosocial factors that could moderate the effects of an intervention. For example, some studies in eco-feedback have found differences in treatment effects based on household income (Vassileva & Campillo, 2014; Podgornik, Sucic & Blazic, 2016), political identity (Costa & Kahn, 2013), or entire countries (Andor et al., 2020). Different self-motives can also affect the effectiveness of feedback interventions by determining how much an individual engages with the feedback through differences in feedback seeking behaviour. Relatedly, in an organisational setting, individuals who have had a long tenure in the organisation were less likely to seek out feedback compared to those who were new (Anseel, Beatty, Shen, Lievens & Sackett, 2015). Differences in goal progress could also influence intervention success. For example, how committed an individual is to a goal determines whether they respond better to positive or negative feedback (Fishbach, Eyal & Finkelstein, 2010). Relatedly, an individual's baseline behaviour may determine how they respond to feedback. In the context of eco-feedback, households with relatively low baseline energy consumption ended up increasing their consumption when presented with social normative feedback (Allcott, 2011).

Finally, individuals may hold different goals related to the behaviour targeted by a feedback intervention. Therefore, individuals may respond differently to different feedback content, depending on the goal that they hold (Gölz & Hahnel, 2016).

3. Research context, aim and structure

This thesis will focus on the context of eco-feedback. Eco-feedback refers to the delivery of feedback to individuals or groups with the goal of reducing their environmental impact. This includes feedback to encourage recycling (Czajkowski, et al. 2019), efficient flying practices (Gosnell et al. 2020), or reducing food waste (Lim et al. 2014). The most popular application of eco-feedback, however, is to reduce household energy and water consumption. There have been a large number of studies that have looked at how providing feedback of consumption to households can encourage reductions in water or energy consumption. The most common design of these interventions is feedback that is compared with descriptive social norms. The first of this design was conceptualised and tested by Nolan et al. (2008) and by Schultz et al. (2007), before being turned into a product by a technology start up, Opower. Since then, it has been widely applied and accepted by utility companies and local authorities all across the world.

There are many benefits of studying feedback in the context of eco-feedback for household water and energy consumption:

1) The behaviour is ubiquitous for almost the entire population, and so the findings are potentially very widely applicable, albeit not immediately generalisable.

2) Household water and energy consumption has a very significant environmental impact. For example, Dietz et al. (2009) calculated that non-regulatory behavioural changes in residential households could potentially save 123 million metric tonnes of carbon per year in the US alone. In contrast to other general areas of feedback or eco-feedback, focusing on this area provides the opportunity to conduct research with the greatest impact.

3) Similarly, with eco-feedback of household water and energy consumption being such a widely applied intervention, there is greater scope of improving its efficacy by identifying the optimal design for these interventions.

4) Household water and energy consumption data is more easily accessible, in that most households have a meter that tracks their water and energy consumption to some degree, and so it is much easier to recruit large samples for experiments.

The aim of the PhD is therefore to better understand how the different components of feedback interventions can affect its overall effectiveness, with the goal of informing optimal designs of feedback interventions. That is, by identifying how much different components may moderate the impact of the intervention, but also how heterogeneity in sample response to the intervention can also affect the impact of the intervention. This aim will be achieved through a series of field experiments that look at different components of eco-feedback interventions. Field experiments were chosen as they are immediately applicable to the real world in which they are tested. Lab experiments may have been more informative in that it is easier to collect a wider range of variables that might better inform the mechanisms by which the effects of the interventions occur. But while these additional variables may be informative, the insights would not be immediately transferable to the real world, as these variables would not be widely available to be used. Once an intervention is scaled up for the wider population, it would become very costly to collect data on those variables. Furthermore, the key insights gained from a field experiment with limited measured variables, is sufficient for informing more impactful designs of eco-feedback interventions.

3.1. Structure and contributions

This thesis is comprised of three papers. Paper 1 is a conceptual replication of eco-feedback interventions. It is the first study of this type of intervention applied to household water consumption in the UK. Not only does it find a 1.8% reduction in water consumption for households that received the treatment, but it also looks at the persistence of these effects after the intervention has ceased. Paper 2 looks at the moderating role of frequency of feedback as well as medium of feedback through two separate field experiments, again in the context of household water consumption in the UK. This paper finds that while delivering frequency through the medium of paper is more effective than by email, frequency did not have a significant moderating effect. Paper 3 explores the unique issue of delivering feedback for both water and energy at the same time in a field experiment conducted in a Middle Eastern country. The aim here was to identify whether combining water and energy into a single metric would be more effective than keeping them separated. Results suggest that while keeping the metrics separated led to larger effects, it was not statistically distinguishable from combining the metrics. This paper also brings important additional insights into how energy eco-feedback interacts with different, non-western, cultures. All three papers also look into the heterogeneity of treatment effects to better understand how targeting or personalisation could help optimise the impact of these interventions.

References

Asimakopoulos, S., Asimakopoulos, G., & Spillers, F. (2016). Motivation and User Engagement in Fitness Tracking: Heuristics for Mobile Healthcare Wearables. Informatics 2017, 4, 5; http://doi:10.3390/informatics4010005

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2007). The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. Journal of Environmental Psychology, 27(4), 265–276. http://doi.org/10.1016/j.jenvp.2007.08.002

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. http://doi.org/10.1016/0749-5978(91)90020-T

Alberts, H. J., Martijn, C., & de Vries, N. K. (2011). Fighting self-control failure: overcoming ego depletion by increasing self-awareness. Journal of Experimental Social Psychology, 47(1), 58e62.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. American Economic Review, 104(10), 3003–3037. http://doi.org/10.1257/aer.104.10.3003

Ammons, R. B. (1956). Effects of Knowledge of Performance: A Survey and Tentative Theoretical Formulation. The Journal of General Psychology, 54(2), 279–299. http://doi.org/10.1080/00221309.1956.9920284

Anseel F, Lievens F, Levy PE. A self-motives perspective on feedback-seeking behaviour: linking organisational behaviour and social psychology research. Int J Manag Rev 2007;9 (3):211–36.

Apte, C., Bibelnieks, E., Natarajan, R., Pednault, E., Tipu, F., Campbell, D., & Nelson, B. (2001). Segmentation-based modeling for advanced targeted marketing. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01 (pp. 408–413). New York, New York, USA: ACM Press. http://doi.org/10.1145/502512.502573

Atkinson, J. W. (1964). An Introduction to Motivation. Oxford, England: Van Nostrand.

Audia, P. G., Locke, E. A., & Smith, K. G. (2000). The paradox of success: An archival and a laboratory study of strategic persistence following radical environmental change. Academy of Management Journal, 43, 837-853.

Balcazar, E, Hopkins, B. L., & Suarez, Y. (1985). A critical, objective review of performance feedback. Journal of Organizational Behavior Management, 7, 65-89.

Balcetis, E. & Cole, S. (2009). "Body in Mind: The Role of Embodied Cognition in Self-Regulation". Social and Personality Psychology Compass. 3 (5): 759–774.

Bandura, A (1977). "Self-efficacy: Toward a Unifying Theory of Behavioral Change". Psychological Review. 84 (2): 191–215. doi:10.1037/0033-295x.84.2.191. PMID 847061.

Bandura, A., & Cervone, D. (1986). Differential engagement of self- reactive influences in cognitive motivation. Organizational Behavior and Human Decision Processes, 38, 92–113.

Bandura, A. (1969). Principles of Behavior Modification. New York: Holt, Rinehart & Winston.

Berkowitz, A. D. (2005). "An overview of the social norms approach". In L. Lederman & L. Stewart (Eds.), Changing the culture of college drinking: A socially situated health communication campaign (193-214). Creskill, New Jersey: Hampton Press.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 6, pp. 1-62). New York: Academic Press.
Brunstein,

Brook, A. (2011). Ecological footprint feedback: Motivating or discouraging? Social Influence, 6(2), 113–128. http://doi.org/10.1080/15534510.2011.566801

Brown, J. D. (1991). Accuracy and bias in self-knowledge. (As cited in Brown, J. D. (1998). The Self. USA: McGraw-Hill)

Burgers, C., Eden, A., Van Engelenburg, M. D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. Computers in Human Behavior, 48, 94–103. http://doi.org/10.1016/j.chb.2015.01.038

Carver, C. S., &Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. Psychological Bulletin, 92, 111–135. http://dx.doi.org/10.1037/0033-2909 .92.1.111

Carver, C. S., & Scheier, M. F. (2012). Cybernetic Control Processes and the Self-Regulation of Behavior. The Oxford Handbook of Human Motivation, (February 2017), 1–30. http://doi.org/10.1093/oxfordhb/9780195399820.013.0003

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 39(5), 752–766.

Champion, M. A., & Lord, R. G. (1982). A control systems conceptualization of the goat-setting and changing process. 0rganizational Behavior and Human Performance, 30, 265-287.

Cialdini, R. 2004. The science of persuasion. Sci. Amer. Mind 284, 76–84.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. Journal of Personality and Social Psychology, 69, 318-328.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The handbook of social psychology (pp. 151-192). New York: McGraw-Hill.

Crommelinck, M., & Anseel, F. (2013). Understanding and encouraging feedback-seeking behaviour: A literature review. Medical Education, 47(3), 232–241. http://doi.org/10.1111/medu.12075

Cunningham, J. A., Hodgins, D. C., Toneatto, T., & Murphy, M. (2012). A Randomized Controlled Trial of a Personalized Feedback Intervention for Problem Gamblers. PLoS ONE, 7(2), e31586. http://doi.org/10.1371/journal.pone.0031586

Czajkowski, M., Zagórska, K., & Hanley, N. (2019) Social norm nudging and preferences for household recycling. Resource and Energy Economics, 58, 101 – 110. https://doi.org/10.1016/j.reseneeco.2019.07.004

Dahling, J. J., Taylor, S. R., Chau, S. L., & Dwight, S. A. (2015). Does coaching matter? A multilevel model linking managerial coaching skill and frequency to sales goal attainment. Personnel Psychology, 69(4), 863–894. http://doi.org/10.1111/peps.12123

Darby, S. (2006). The effectiveness of feedback on energy consumption: A Review for DEFRA of the literature on metering, billing and direct displays. Oxford, UK: Environmental Change Institute.

de Bruin, M., Sheeran, P., Kok, G., Hiemstra, A., Prins, J. M., Hospers, H. J., & van Breukelen, G. J. P. (2012). Self-regulatory processes mediate the intention-behavior relation for adherence and exercise behaviors. Health Psychology, 31, 695–703. http://dx.doi.org/10.1037/ a0027425

DePasque, S., & Tricomi, E. (2015). Effects of intrinsic motivation on feedback processing during learning. http://doi.org/10.1016/j.neuroimage.2015.06.046

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. The Psychological Record, 54, 207–231.

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). MINDSPACE: Influencing Behaviour through Public Policy. Retrieved September 27, 2016, from: http://www.instituteforgovernment.org.uk/publications/mindspace

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. Journal of Economic Psychology, 47, 1–16. http://doi.org/10.1016/j.joep.2014.12.003

Dossett, D. L., Latham, G. P., & Mitchell, T. R. (1979). Effects of assigned vs. participatively set goals, knowledge of results, and individual differences on employee behavior when goal difficulty is held constant. Journal of Applied Psychology, 64, 291–298.p

Durham, C., Knight, D., & Locke, E. (1997). Effects of leader role, team-set goal difficulty, efficacy, and tactics on team effectiveness. Organizational Behavior and Human Decision Processes, 72, 203–231.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Row Peterson

Foundation, E. E. (2013). Sutton Trust - Education Endowment Foundation Teaching and Learning Toolkit: Technical Appendices. Retrieved January 29, 2018, from https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? Energy Efficiency, 1(1), 79–104. http://doi.org/10.1007/s12053-008-9009-7

Fishbach, A., Dhar, R., & Zhang, Y. (2006). Subgoals as substitutes or complements: the role of goal accessibility. Journal of Personality and Social Psychology, 91(2), 232–242. http://doi.org/10.1037/0022-3514.91.2.232

Fishbach, A., & Finkelstein, S. R. (2012). How Feedback Influences Persistence, Disengagement, and Change in Goal Pursuit. In H. Aarts & E. Aronson (Eds.), Goal-directed behavior (pp. 1–53). Psychology Press. http://doi.org/10.4324/9780203869666

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA: Addison Wesley.

Fox, S., & Duggan, M. (2012). Mobile health 2012. Washington DC: Pew Research Center's Internet & American Life Project. www. pewinternet.org/Reports/2012/Mobile-Health.aspx.

Fujita, K., & Carnevale, J. J. (2012). Transcending Temptation Through Abstraction: The Role of Construal Level in Self-Control. Current Directions in Psychological Science, 21(4), 248–252. http://doi.org/10.1177/0963721412449169

Fuller, C., Michie, S., Savage, J., McAteer, J., Besser, S., Charlett, A., … Stone, S. (2012). The Feedback Intervention Trial (FIT) - Improving Hand-Hygiene Compliance in UK Healthcare

Workers: A Stepped Wedge Cluster Randomised Controlled Trial. PLoS ONE, 7(10), e41617. http://doi.org/10.1371/journal.pone.0041617

García-González, L., Moreno, M. P., Moreno, A., Gil, A., & Del Villar, F. (2013). Effectiveness of a video-feedback and questioning programme to develop cognitive expertise in sport. PLoS ONE, 8(12), e82270

Gerber, A. S., & Rogers, T. (2009). Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You. The Journal of Politics, 71(1), 178–191. http://doi.org/10.1017/S0022381608090117

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology, 38, 69–120. http://dx.doi.org/10.1016/ S0065-2601(06)38002-1

Gosnell, G.K., List, J.A., & Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. Journal of Political Economy 128(4). 1195-1233

Harmon-Jones, E., & Mills, J. (Eds.). (1999). Cognitive dissonance: Progress on a pivotal theory in social psychology. Washington, DC: American Psychological Association.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 16–7. http://doi.org/10.3102/003465430298487

Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., … Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. Psychological Bulletin, 142(2), 198–229. http://doi.org/10.1037/bul0000025

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. Perspectives on Psychological Science, 11(6), 838–854. http://doi.org/10.1177/1745691616650285

Hanna, G. S. (1976). Effects of total and partial feedback in multiple-choice testing upon learning. The Journal of Educational Research, 69, 202–205.

Hermsen, S., Frost, J., Renes, R. J., & Kerkhof, P. (2016). Using feedback through digital technology to disrupt and change habitual behavior: A critical review of current literature. Computers in Human Behavior. http://doi.org/10.1016/j.chb.2015.12.023

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as Reference Points. Cognitive Psychology, 38(1), 79–109. http://doi.org/10.1006/cogp.1998.0708

Hoggan, E., & Brewster, S. A. (2010). Crosstrainer: testing the use of multimodal interfaces in situ. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 333e342). ACM

Hughes, G. (2011). Towards a personal best: a case for introducing ipsative assessment in higher education. Studies in Higher Education, 36(3), 353–367. http://doi.org/10.1080/03075079.2010.486859

Hughes, G., Wood, E., & Kitagawa, K. (2014). Use of self-referential (ipsative) feedback to motivate and guide distance learners. Open Learning: The Journal of Open, Distance and E-Learning, 29(1), 31–44. http://doi.org/10.1080/02680513.2014.921612

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organization. Journal of Applied Psychology, 64, 349-371.

Jones, G. M. (1910). Experiments on the reproduction of distance as influenced by suggestions of ability and inability. Psychological Review, 17, 269–278. http://dx.doi.org/10.1037/h0074504

Judd, C. H. (1905). Practice without knowledge of results. The Psychological Review: Monograph Supplements, 2, 185–198.

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. Psychological Bulletin, 141(6), 1205–1227. http://doi.org/10.1037/a0039650

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. Journal of Marketing Research, 43(1), 39-58.

Kluger, A. N., & DeNisi, A. (1998). Feedback Interventions: Toward the Understanding of a Double-Edged Sword. Current Directions in Psychological Science. http://doi.org/10.1111/1467-8721.ep10772989

Krajcsi, A., & Kojouharova, P. (2017). Symbolic Numerical Distance Effect Does Not Reflect the Difference between Numbers. Frontiers in Psychology, 8, 2013. http://doi.org/10.3389/fpsyg.2017.02013

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 34, pp. 331– 378). San Diego, CA: Academic Press.

Lally, P., & Gardner, B. (2013). Promoting habit formation. Health Psychology Review, 7(sup1), 137e158.

Latham, G. P. (2001). The importance of understanding and changing employee outcome expectancies for gaining commitment to an organizational goal. Personnel Psychology, 54, 707–716.

Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: challenging the "more is better" assumption. Organizational Behavior and Human Decision Processes, 116(2), 217e228.

Li, M., & Chapman, G. B. (2013). A big fish or a small pond? Framing effects in percentages. Organizational Behavior and Human Decision Processes, 122(2), 190–199. http://doi.org/10.1016/J.OBHDP.2013.07.003

Locke, E. (1991). Goal theory vs. control theory: Contrasting approaches to understanding work motivation. Motivation and Emotion, 15(1).

Mangiapanello, K. A., & Hemmes, N. S. (2015). An Analysis of Feedback from a Behavior Analytic Perspective. The Behavior Analyst, 38(1), 51–75. http://doi.org/10.1007/s40614-014-0026-x

Nauges, C., & Whittington, D. (2019). Social Norms Information Treatments in the Municipal Water Supply Sector: Some New Insights on Benefits and Costs. Water Economics and Policy, 5(3), **https://doi.org/10.1142/S2382624X18500261**

Nisa, C., Belanger J.J., Schumpe, B.M., & Faller, D.G. (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. Nature Communications, 10.

Jones, G. M. (1910). Experiments on the reproduction of distance as influenced by suggestions of ability and inability. Psychological Review, 17, 269–278. http://dx.doi.org/10.1037/h0074504

Judd, C. H. (1905). Practice without knowledge of results. The Psychological Review: Monograph Supplements, 2, 185–198.

Klein, H. J., Wesson, M. J., Hollenbeck, J. R., & Alge, B. J. (1999). Goal commitment and the goal-setting process: Conceptual clarification and empirical synthesis. Journal of Applied Psychology.

Kruger, J., & Vargas, P. (2008). Consumer confusion of percent differences. Journal of Consumer Psychology, 18(1), 49–61. http://doi.org/10.1016/J.JCPS.2007.10.009

Li, Y., & Guo, Y. (2016). Wiki-Health: From Quantified Self to Self-Understanding. Future Generation Computer Systems, 56, 333–359. http://doi.org/10.1016/j.future.2015.08.008

Liberman, N., & Förster, J. (2008). Expectancy, value and psychological distance: A new look at goal gradients. Social Cognition, 26(5), 515–533.

Locke, E. A., Chah, D., Harrison, S., & Lustgarten, N. (1989). Separating the effects of goal specificity from goal level. Organizational Behavior and

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705–717. http://doi.org/10.1037//0003-066X.57.9.705

Lupton, D. (2016). The Quantified Self. Foundations of Augmented Cognition, 1–237. http://doi.org/10.1089/big.2012.0002

MacKay, D. M. (1966). Cerebral organization and the conscious control of action. In J. C. Eccles (Ed.), Brain and conscious experience (pp. 422–445). Berlin: Springer-Verlag.

Matsui, T., Okada, A., & Mizuguchi, R. (1981). Expectancy theory predictions of the goal theory postulate, "The harder the goals, the higher the performance." Journal of Applied Psychology, 66, 54–58.

Mattheiss, S. R., Alexander, E. J., & Graves, W. W. (2017). Elaborative feedback: Engaging reward and task-relevant brain regions promotes learning in pseudoword reading aloud. Cognitive, Affective and Behavioral Neuroscience, 1–20. http://doi.org/10.3758/s13415-017-0553-5

McCalley, L. T., &Midden, C. J. H. (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social- orientation. Journal of Economic Psychology, 23, 589–603.

McCambridge, J., Bendtsen, M., Karlsson, N., White, I. R., Nilsen, P., & Bendtsen, P. (2013). Alcohol assessment and feedback by email for university students: Main findings from randomised controlled trial. The British Journal of Psychiatry, 203, 334–340. http://dx.doi.org/10.1192/bjp.bp.113.128660.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: A meta-regression. Health Psychology, 28, 690–701. http://dx.doi .org/10.1037/a0016136

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the structure of behavior. New York: Holt, Rinehart, & Winston.

Muralidharan, K., Singh, A., & Ganimian, A. J. (2016). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. NBER Working Paper Series, 85. http://doi.org/10.3386/w22923

Nisa, C., Belanger J.J., Schumpe, B.M., & Faller, D.G. (2019). **Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. Nature Communications, 10.**

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative social influence is underdetected. Personality & Social Psychology Bulletin, 34(7), 913–23. http://doi.org/10.1177/0146167208316691

Opitz, B., Ferdinand, N. K., & Mecklinger, A. (2011). Timing matters: The impact of immediate and delayed feedback on artificial language learning. Frontiers in Human Neuroscience, 5, 8.

Orehek, E., Bessarabova, E., Chen, X., & Kruglanski, A. W. (2011). Positive affect as informational feedback in goal pursuit. Motivation and Emotion, 35(1), 44–51. http://doi.org/10.1007/s11031-010-9197-2

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. Journal of Applied Social Psychology, 34, 243–281.

Powers, W. T. (1973). Behavior: The control of perception. Chicago, IL: Aldine. Reyna,

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55, 68-78.

Salmoni, A. W, Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. Psycho- logical Bulletin, 95, 355-386.

Sanguinetti, A., Dombrovski, K., & Sikand, S. (2017). Information, timing, and display: A design-behavior framework for improving the effectiveness of eco-feedback. http://doi.org/10.1016/j.erss.2017.10.001

Sedikides C, Strube MJ. Self-evaluation: to thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. Adv Exp Soc Psychol 1997; 29:209–69.

Service, O., Hallsworth, M., Halpern, D., Algate, F., Gallagher, R., Nguyen, S., … Kirkman, E. (2014). EAST Four simple ways to apply behavioural insights. Nesta, 53. http://doi.org/http://behaviouralinsights.co.uk/publications/east-four-simple-ways-apply-behavioural-insights

Siero, F. W., Bakker, A. B., Dekker, G. B., & van den Burg, M. T. (1996). Changing organizational energy consumption behavior through comparative feedback. Journal of Environmental Psychology, 16, 235–246. http://dx.doi.org/10.1006/ jevp.1996.0019.

Schultz, P. W. (1999). Changing behavior with normative feedback interventions: A field experiment on curbside recycling. Basic and Applied Social Psychology, 21, 25–36. http://dx.doi.org/10.1207/15324839951036533

Schunk, D. H., & Swartz, C. W. (1993). Goals and Progress Feedback: Effects on Self-Efficacy and Writing Achievement. Contemporary Educational Psychology, 18(150), 337–354. http://doi.org/10.1006/ceps.1993.1024

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. European Review of Social Psychology, 12, 1–36. http://dx.doi.org/10.1080/14792772143000003

Sheeran, P., & Webb, T. L. (2011). From goals to action. In H. Aarts & A. Elliott (Eds.), Frontiers in social psychology: Goal-directed behavior (pp. 175–202). London, United Kingdom: Psychology Press.

Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. New York, NY: Appleton-Century.

Swann, W. B., Jr. (1990). To be adored or to be known? The interplay of self-enhancement and self-verification. (As cited in Brown, J. D. (1998). The Self. USA: McGraw-Hill)

Tao, T., Wyer, R. S., Jr., & Zheng, Y. (2017). The role of categorization and scale endpoint comparisons in numerical information processing: A two-process model. Journal of Experimental Psychology: General, 146(3), 409-427.

Thorndike, E. L. (1927). The law of effect. The American Journal of Psychology, 39, 212–222. http://dx.doi.org/10.2307/1415413

Tolli, A. P., & Schmidt, A. M. (2008). The role of feedback, casual attributions, and self-efficacy in goal revision. The Journal of Applied Psychology, 93(3), 692–701. http://doi.org/10.1037/0021-9010.93.3.692

Triandis, H. C. (1977). Interpersonal behavior. Monterey, CA: Brooks/ Cole.

VandeWalle, D. (2003). A goal orientation model of feedback-seeking behavior. Human Resource Management Review, 13(4), 581–604. http://doi.org/10.1016/j.hrmr.2003.11.004

Vassileva, I., Odlare, M., Wallin, F., & Dahlquist, E. (2012) The impact of consumers' feedback preferences on domestic electricity consumption, Appl. Energy 93, 575–582

Vollmer, A.-L., Mühlig, M., Steil, J. J., Pitsch, K., Fritsch, J., Rohlfing, K. J., & Wrede, B. (2014). Robots Show Us How to Teach Them: Feedback from Robots Shapes Tutoring Behavior during Action Learning. PLoS ONE, 9(3), e91349. http://doi.org/10.1371/journal.pone.0091349

Wallace, S. G. & Etkin, J. (2018). How Goal Specificity Shapes Motivation: A Reference Points Perspective, Journal of Consumer Research, 44(5), Pages 1033–1051

Webb, T. L., Chang, B. P. I., & Benn, Y. (2013). "The Ostrich Problem": Motivated avoidance or rejection of information about goal progress. Social and Personality Psychology Compass, 7(11), 794–807. http://doi.org/10.1111/spc3.12071

Wright P. M., Kacmar K. M. (1994), "Goal Specificity as a Determinant of Goal Commitment and Goal Change," Organizational Behavior and Human Decision Processes, 59 2, 242–60.

Zahiri, M., Nelson, C. A., Oleynikov, D., & Siu, K. C. (2017). Evaluation of Augmented Reality Feedback in Surgical Training Environment. Surgical Innovation, 1, 155335061773942.

4. Paper 1: Social norms based eco-feedback for household water consumption

This paper was published in Sustainability 13(5) in March 2021:
https://doi.org/10.3390/su13052796

Abstract: Physical water scarcity is a growing threat to the lives of everyone in the world. Non-pecuniary interventions that encourage water conservation amongst households are an effective tool to promote sustainable consumption. In a randomised field experiment on 3,461 UK households, a social norms based eco-feedback intervention was found to reduce water consumption by around 5.43 litres a day or by 1.8% over 29 months. This effect did not persist for the 10 months after the intervention was stopped suggesting a lack of habit formation. Households with low consumption at baseline reduced their consumption the most. Heterogeneity was also found across quantile treatment effects, where households in the top and bottom quantiles increased their consumption. The results further contribute to the growing evidence on the effectiveness of combining social norms and eco-feedback as an intervention for conservation.

4.1. Introduction

Physical water scarcity is a growing global problem that affects people across every continent. In 2018, Cape Town became the first major city to run out of water, and there are likely many more to follow ('The 11 Cities Most Likely to Run out of Drinking Water - like Cape Town', 2018). While a lack of rainfall due to climate change's effects plays a key role in these droughts, residential water demand increases is also a major contributor. In the past 100 years, global water demand has increased by 600% (Wada et al., 2016). For example, in the UK, per person consumption has increased from around 85 litres per day in the 1960s to around 140 litres per day in 2016 (Rob Lawson et al., 2018). Globally, household demand for water increases as populations and economies continue to grow (Boretti & Rosa, 2019; Hussien et al., 2016).

To curtail this growing demand, utilities and governments have a range of policies and interventions to choose from to motivate households to reduce their water consumption. Pecuniary policies, such as water price increase, increasing block tariffs, or peak-time pricing, while common, have not been shown to be very effective at reducing water demand, potentially due to water demand's price inelasticity (Reynaud & Romano, 2018). On the other hand, non-pecuniary based interventions that try to encourage pro-environmental behaviours by leveraging intrinsic motivation are a more promising approach (Lehner et al., 2016). An intervention that has continued to grow in popularity due to its effectiveness is eco-feedback.

Eco-feedback refers to providing information to people regarding their environmental behaviours, to reinforce and/or encourage behaviour change. Leveraging on the concept of self-regulation and cybernetics, feedback works by encouraging people to compare their own behaviours against a set standard (Carver & Scheier, 2012; Kluger & DeNisi, 1996). When a discrepancy is detected, people become motivated to decrease the gap between their behaviour and the standard (Buchanan & Russo, 2015). Additionally, eco-feedback can play a role in making consumption more visible. Within the literature, eco-feedback has been framed with no comparison, comparison against historical consumption, the consumption cost, comparison against a set goal or comparison against a social norm (Abrahamse et al., 2005; Fischer, 2008; Karlin et al., 2015). Of these, comparisons against social norms have received a lot of attention in both research and policy settings.

Social norms refer to the informal understanding individuals have about what most other people do or approve of (Cialdini & Trost, 1998; Schultz et al., 2007). Social norm messaging has been shown to be an effective tool in influencing behaviour in various contexts (Bicchieri & Dimant, 2019; Cialdini & Trost, 1998; Czajkowski et al., 2019; Kandul et al., 2020; Lehner et al., 2016; Mahler et al., 2008; Perkins, 2002). This could be because individuals refer to others' behaviours to help guide their own actions (Sherif, 1936). While early forms of social norms based interventions typically just provide descriptive information regarding the behaviour of others (e.g., "80% of hotel guests reuse their towels"), more recent applications combine this social information with a comparison of the individual's behaviour (e.g., "Your peers go to the gym twice as much as you do"). This not only highlights what is normative behaviour, as dictated by social peers, but also directly highlights how far an individual's behaviour is from the social norm. Therefore, social norms are used by individuals to first create a reference point to make sense of their consumption, and second, to ensure that their behaviours align with what is socially acceptable based on peer behaviour (Schultz et al., 2007).

Social norms based eco-feedback was first conceptualised and tested in a field setting by Nolan et al. (Nolan et al., 2008) and Schulz et. al (2007). They looked at the effectiveness of descriptive social norms messaging on reducing household energy use, compared to other types of messaging such as pro-environmental, financial savings, or social responsibility. This approach was then further developed by a software company, Opower, and tested on a much larger scale by Allcott (Allcott, 2011). The intervention was found to reduce household energy consumption by around 2.2%. The intervention have since been conceptually replicated across other countries but was found to have smaller effect sizes in Germany (Andor et al., 2020) and Italy (Bonan et al., 2020). Similarly, this approach has also effectively reduced household water consumption by around 4% in the US and Colombia (Bhanot, 2017, 2018; Brent et al., 2015; P. Ferraro & Price, 2011; Jaime Torres & Carlsson, 2018; Schultz et al., 2016).

The design of these interventions typically involves displaying, through a graph, a household's energy/water consumption compared to the average energy/water consumption of surrounding households. The inclusion of descriptive social norms messaging alone works with high consuming households but can potentially cause a 'boomerang' effect among low-consuming households. This is because the message that they are performing better than their peers creates a moral license for them to consume more (Schultz et al., 2007). To counter this, an injunctive norm message, which highlights a standard of behaviour that the social group approves, is combined with a descriptive norm message to reinforce low-consuming households' efficient behaviour while still encouraging the behaviour change of high-consuming households [15, 31]. The intervention materials are known as home energy/water reports in the industry.

4.2. Long term effects

One area of interest in these studies is understanding how lasting the treatment effects are after the intervention has ceased. Ferraro and Miranda (P. J. Ferraro & Miranda, 2013) found that the effects of sending one instance of a social norms based eco-feedback intervention reduced household water consumption over the subsequent two years, but this effect decayed year on year. Having delivered the intervention in May 2007, the initial effect in the following summer was -1.74%. By the summer of 2008, it had decreased to -0.64%, and then to -0.35% by the summer of 2009.

One potential way of overcoming this decay in the effects is through repeated exposure to the intervention. In Allcott's (2011) field experiments on household energy consumption, the social norms based eco-feedback intervention was delivered repeatedly, on a monthly or quarterly basis, for about four years. While a decay in the effect was observed between each home energy report, each new report would renew the intervention's effect. This cycle of backsliding and renewing of the effect attenuated over repeated exposure of the intervention. Finally, after two years of the intervention, a random selection of households in the treatment group stopped receiving the reports. There was no longer a decay of the treatment effect for those households at the same rate, suggesting that households had formed a habit of energy conservation.

4.3. Heterogeneous treatment effects

Another important consideration in studying social norms based eco-feedback interventions is the heterogeneity of its treatment effects. While factors such as a household's demographic or psychographic characteristics may lead to differences in the impact an intervention might have, a primary factor driving heterogeneity in these types of interventions is pre-treatment consumption (Allcott, 2011; Czajkowski et al., 2019). By running regressions that interact the treatment effect with deciles of pre-treatment consumption, conditional average treatment effects (CATE) can be observed. Across various studies in energy and water, households in the top deciles have higher effect sizes, suggesting that high consumers in the baseline period were more likely to be affected by the treatment (Allcott, 2011; Brent et al., 2015; Jaime Torres & Carlsson, 2018).

Another approach to identifying heterogeneity in treatment effects is using Quantile Treatment Effects (QTEs). These are differences in consumption between the treatment and control groups on each corresponding quantile across the distribution of treatment effects. Allcott (Allcott, 2011) found heterogeneity in the treatment effect, with the effect being stronger in the upper quantiles. Ferraro and Miranda (P. J. Ferraro & Miranda, 2013) find a similar pattern of heterogeneity following the first year of post-treatment. Taken together, Allcott argues that this suggests that the 'boomerang effect' may not be as strong as expected, possibly due to the presence of the injunctive norm in these interventions.

4.4. Aim

The current study aims to identify and understand the treatment effects of a social norms based eco-feedback intervention on household water consumption. While not a direct replication, the current study applies a very similar intervention and experimental approach to previous research. This study will be first the intervention of its kind to be applied to household water consumption in the UK. The UK is an important area of study because while the majority of survey respondents believe that the UK is a 'wet and rainy' country with an abundance of water, the reality is that water scarcity is a reality and an ever growing problem (DEFRA, 2020). This means that a social norms based intervention may be more effective than traditional water conservation campaigns, especially when most people may not believe water conservation to be a pressing issue. This is partly why many water utilities in the UK who have run water conservation campaigns are interested in this form of intervention for its customers. This study is also the first study to look at the long term effects of this type of repeated intervention on water consumption. While, Ferraro et. al (P. J. Ferraro et al., 2011), did study the long term effects of their intervention, they only had one instance of the treatment delivered, which may not have been enough to form a long term habit in the same vein as Allcott's study.

The main hypothesis is that households in the treatment group, that receive the social norms based eco-feedback intervention, will reduce their consumption more than households in the control group. This effect is also expected to be greater for households with higher consumption levels at baseline. The long term effects of the intervention and the heterogeneity of the treatment effects will also be explored.

4.5. Materials and Methods

4.5.1. Participants and design

Participants for this study are made up of residential customers from South East Water (SEW), a water utility in the UK. SEW services a water-stressed area of the UK with a growing population and a higher than average water demand. The average consumption of households in the region is 157 litres per person per day. Customers are billed for their usage on a six-monthly basis at different intervals based on their geographic location. 4,000 households were randomly selected from SEW's entire customer base and were opted into the programme as part of the SEW's water efficiency strategy. These 4,000 households were then randomly allocated on the household level into a treatment and control group. Those in the treatment group would receive the eco-feedback intervention, and the control group did not receive anything.

Commencing in November 2016, the intervention was delivered at roughly 6-monthly intervals. The dates and the number of households that received the treatment at each interval are presented in table 1. Those dates represent the day the reports were generated. Once generated, they were passed to a third-party that printed and posted the reports to households. The length of time between when the reports were generated and when households received the report was roughly one week. At each interval, households were excluded from the programme entirely or from receiving the intervention at that interval based on a set of criteria. The exclusion criteria were 1) households that had moved out of their homes or became deceased, 2) households that are on a social tariff that provides special rates for people on a low income or those that have specific medical conditions, 3) households that opted out of the programme, 4) households that had unusual consumption of above 10,000 litres a day, because these are very unlikely to be households but rather a commercial building, and 5) households that were consuming 100% more than their average neighbours. The latter was included to help reduce any potential customer complaints for SEW. Not all households in the treatment group received the intervention at each interval due to the listed criteria for exclusions, their consumption amount was still tracked and recorded as long as they had not closed their account. The last intervention was delivered to the treatment group in March 2019, but consumption data continued to be collected until February 2020. Finally, after one year of the programme, a phone survey was conducted, and 200 households were interviewed to measure recall and perception of the programme. The survey was not specifically for this programme, but a few questions were added to a wider survey conducted regularly by SEW. There was not an option to conduct a pre-intervention survey.

Date when treatment was generated	Number of households included
18th November 2016	1,949
16th of June 2017	1,768
9th of December 2017	1,559
30th of July 2018	1,277
19th of February 2019	1,035

Table 1 Dates of when the home water reports were generated, and the number of households at each of those dates.

4.5.2. Materials

Households in the treatment condition received a paper home report and were given access to an online portal that contained the same content of the home report. The paper report and online portal were developed by Advizzo, a software-as-a-service company that works with utilities. The home report's main feature was the social norms messaging in the form of a neighbour comparison graph. The neighbour comparison graph displays a comparison of the households' water consumption against the mean consumption of households of similar occupancy within their area and the top 20% most efficient households. Households are placed into one of three groups: 'more than average' – those consuming more than the mean of similar households, 'below the average' – those consuming less than the mean of similar households and 'most efficient' – those in the top 20% of households with the least consumption. This comparison is displayed using a bar graph where the bars represent the amount consumed. Consumption is displayed as cubic litres (m^3) to align with what households see in their water bills. This neighbour comparison graph serves to deliver descriptive social norms and has been used widely in previous research (Allcott, 2011; Ayres et al., 2013; Schultz et al., 2016). In addition to this, an injunctive social norm is also displayed in the form of a series of 'smiley faces' with three labelled levels, 'more than average', 'Good' and 'Great' to counter any possible boomerang effect. Finally, the home report also included three tips on ways to reduce water consumption. Figure 1 displays an example of the home report.

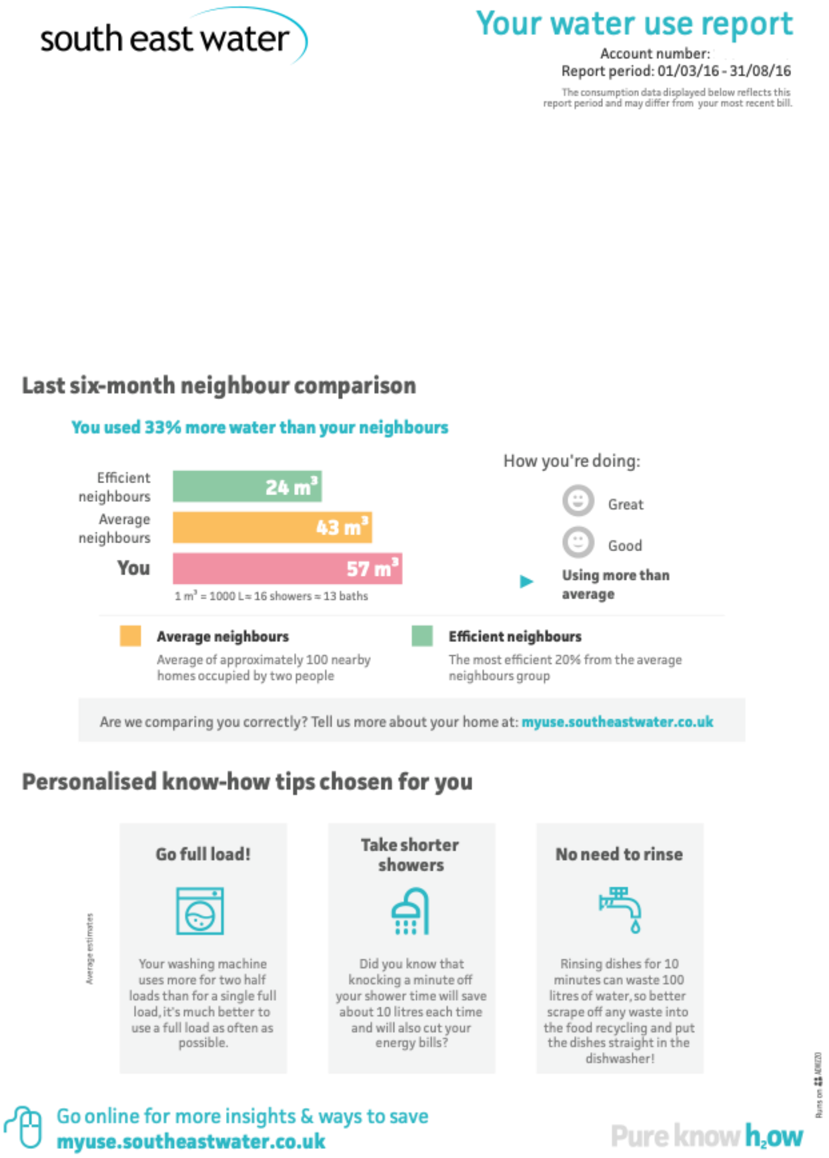

Figure 1 An example of the home report sent to households in the treatment group.

4.5.3. Data

Consumption data were collected through water meters installed on the household level. These are read on a 6-monthly basis. Each house has its meter read at different times depending on where the house is in respect to the meter reader's route at a given period. Water consumption is determined by looking at the difference in time between a previous meter read and the most recent meter read to create a cubic meter volume of water for that period. That volume is then divided by the number of months in that period to create a monthly consumption amount. This is because the data is stored as monthly intervals by Advizzo. The monthly consumption is then finally converted to daily consumption values by dividing the monthly reads by the number of days in each respective month. Since the treatment is delivered as paper mail, it is very difficult to know exactly when a household decides to read the report and be treated. Therefore, the difference between when a household is treated and when their meter is read is not accounted for in the estimations below. Furthermore, South East Water's billing intervals were not provided, so I could not control for any potential billing shocks that could affect water consumption.

183 households that either did not have any baseline consumption (n=182) or any post-treatment consumption (n=1) were dropped from the dataset. Aside from this, if a household had received at least one report, their consumption data were kept within the analysed dataset. At the point of analysis 3,461 households were included in the final data set, with 1,745 households in the control group and 1,716 households in the treatment group. 65.7% of those households had data available across all 40 months, with about half of those in the treatment group. Not having all 40 months is defined as attrition. A regression of the number of months of data available per household on treatment assignment showed no statistically significant effect (p=0.374). Additionally, a regression of a binary indicator of attrition on treatment assignment showed no statistically significant effect (p=0.890). These two analyses to test the balance of attrition, and the removal of the 183 households without baseline or post-treatment data were not part of the analysis plan and were only conducted post-hoc. Finally, 336 observations that fell below and above the 1st and 99th percentiles were treated as outliers and were dropped from the dataset.

Mean consumption over the one year before the programme commenced between the treatment (M=285.59 litres/day, SD=141.15) and control (M=279.09 litres/day, SD=142.42) groups were not statistically significantly different (p=0.079). This was estimated using a regression with the same specification used to estimate the average treatment effect, but only with data in the pre-treatment period. Furthermore, the coefficients in the pre-treatment period of Figure 2 are very close to zero, suggesting that the randomisation resulted in balanced experimental groups. Average consumption during the 40 months post-intervention period for the treatment group was (M=294.00 litres/day, SD=164.98), and for the control group was (M=291.49, SD=162.87). A table of descriptive statistics in the appendix shows mean consumption for each period alongside sample size for each group.

4.6. Results

4.6.1. Average treatment effects treatment period

To estimate the average treatment effects of the intervention across the first 29 months of the programme, I use an ANCOVA model similar to that of Allcott (Allcott & Rogers, 2014).

$$LPD_{it} = T_i + LPDpre_{it} + \pi_m + \epsilon (1)$$

Where LPD_{it} is household $i's$ water consumption in litres per day on month t. T_i is the treatment indicator. $LPDpre_{it}$ is average water consumption in the matching calendar month in the baseline period. π_m are month and year fixed effects. Only data from the 29 months following the first treatment letter were included in this analysis. Standard errors are robust and clustered to the household level to control for autocorrelation.

Results indicate that during the first 29 months of the intervention there was a significant main effect for the treatment group, F (16,90644) = 159.6, p<0.05. Households in the treatment group reduced their consumption by 5.43 litres per day on average compared to the control group. This is a 1.8% reduction from households' mean daily consumption in the control group. For perspective, this is equivalent to leaving the tap running for a minute.

Alternatively, a difference-in-difference model with the following specification could also be used, where P_{it} represents the post-treatment indicator and data from both the pre and post intervention period were included:

$$LPD_{it} = T_i P_{it} + \pi_m + \epsilon \qquad (2)$$

Results comparing the two models is available in Table 2. The difference-in-difference model suggests that consumption of the treatment group was not statistically significantly different from that of the control group. While a difference-in-difference model could directly control for time-invariant omitted variables (such as sociodemographics and property characteristics), this is not a primary concern here because this study randomises households into the treatment groups which already guarantees that in expectation such variables are not correlated with treatment assignment. In addition, the ANCOVA model used here has the benefit of more power and a higher R-squared, per the results in Table 2 and in line with the econometric analysis of McKenzie (2012).

	ANCOVA	Difference-in-difference
Treatment	-5.430**	-5.147
	(2.429)	(4.687)
Baseline	Yes	No
Time FE	Yes	Yes
Number of observations	90,661	134,484
R2	0.479	0.002

Table 2 Regression results of the average treatment effect for both ANCOVA and difference-in-difference models. *p<0.1, **p<0.05, ***p<0.001

Figure 2 of an event study graph shows the average treatment effect across each month of the entire programme based on the model above. This was calculated by running the regression of Equation (1) with an interaction term between group assignment and month (Equation 3). Figure 2 demonstrates the fluctuation in the treatment effect across the programme period, with the greatest reductions occurring in the first summer months. Furthermore, the average treatment effect appears to not be as strong as in the first year, with greater increases in consumption being observed until the end of the programme. It should be noted that the treatment effect is mostly just slowing down what is essentially an increase in consumption across all households. Households in the control group increased its consumption by 12.25 litres per day between the pre and post treatment period, while the experimental group only increased its consumption by 5.06 litres per day.

$$LPD_{it} = T_i\pi_m + LPDpre_{it} + \epsilon \qquad (3)$$

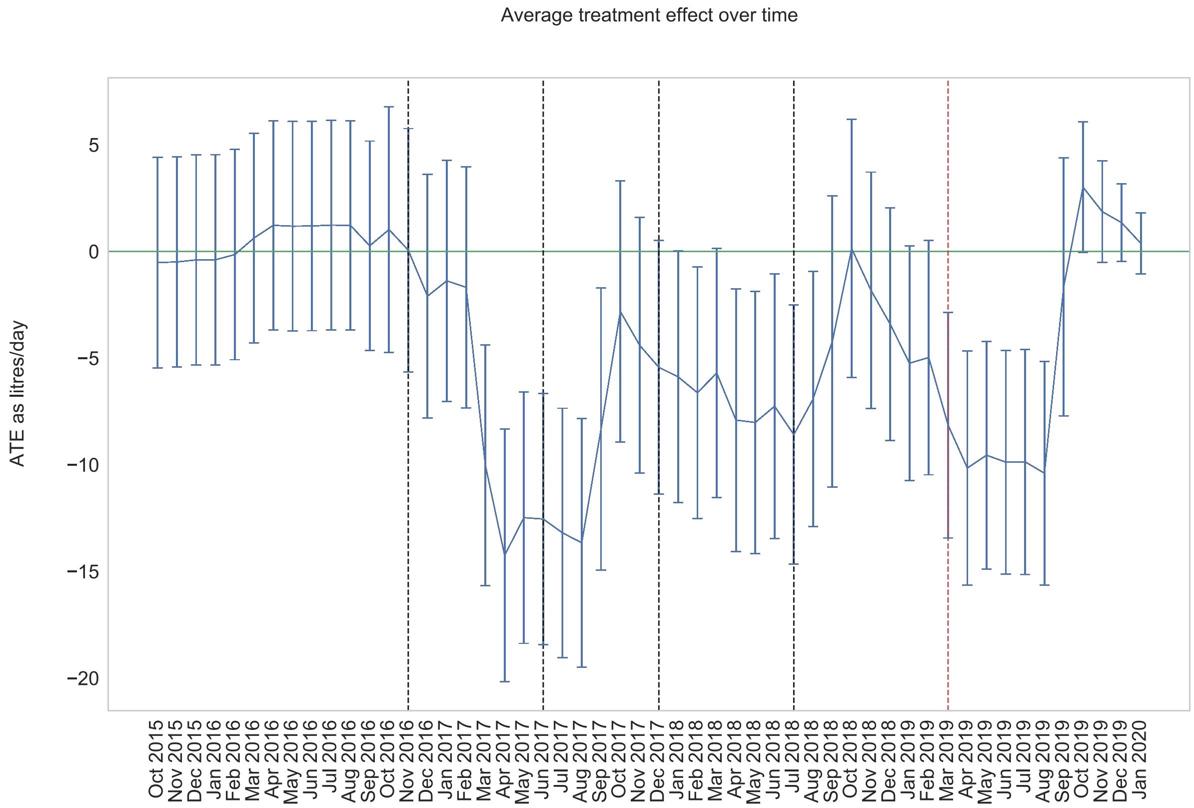

Figure 2 Event study graph of changes in the intervention's average treatment effect in litres per day across all months of the pre and post-intervention period compared to the control group. The coefficients and standard error were generated from interacting treatment assignment with monthly dummy variables while controlling for baseline consumption. Each vertical dashed line indicates when the intervention was administered. Note: the version of this graph in Appendix A1 uses the dataset that does not remove outliers above and below the 1st and 99th percentiles and shows a stronger upward trend after month 29.

4.6.2. Quantile treatment effects

Quantile treatment effects for the first 29 months of the programme can be seen in Figure 3. This was estimated as an unconditional quantile treatment effect with only treatment assignment as the independent variable, and household water consumption in litres per day as the dependent variable (Equation 4). The standard errors were clustered on the household level. Households in the bottom 20% seem to be increasing their consumption, suggesting a potential boomerang effect whereby being labelled "Efficient Neighbours" may have licensed them to increase their consumption. The greatest reduction in consumption appears to occur across the middle distribution. There also appears to be some households that are notably increasing their consumption as evidenced by the mass moving out to the upper tail. This could be households that have been labelled as high consumers giving up on being efficient. Without a rank invariance assumption, however, these statements are only suggestive. For quantiles 0.5 to 0.95 in increments of 0.1, the following model is fit to each quantile separately, where τ is a representative quantile, and the quantile is an inverse cumulative distribution function. And T_i denotes treatment group assignment.

$$Q_{lpd|Ti}(\tau) = \beta_0 + \beta_{Treatment} T_i \qquad (4)$$

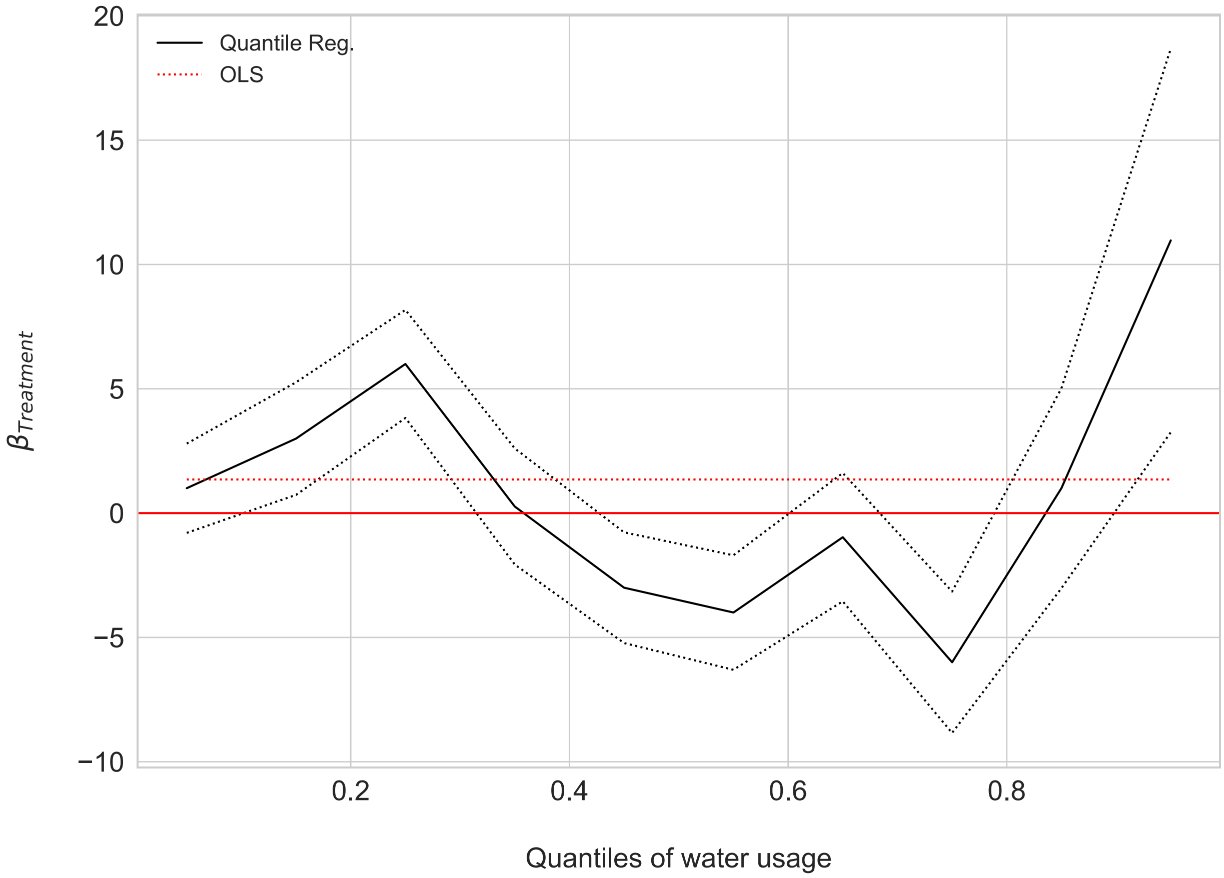

Figure 3 Quantile treatment effects for the treatment across the first 29 months of the intervention. The dotted red line represents the unconditional average treatment effect. It shows an increase in consumption because it has not controlled for baseline consumption in the same way the main model specification (equation 1) for the ATE has. The black line represents the unconditional quantile treatment effect. Black dotted lines represent the 95% confidence intervals. Standard errors were clustered on the household level.

Figure 4 shows quantile treatment effects for the 10 months after the final treatment was delivered. As a purely exploratory interpretation, it appears as though those in the bottom quantiles are increasing their consumption at a slower rate, while those in the top quantiles are increasing their consumption at a faster rate. Overall, the similarity in shape between Figure 3 and Figure 4 suggests that the decay of treatment effect occurs relatively similarly across most of the distribution.

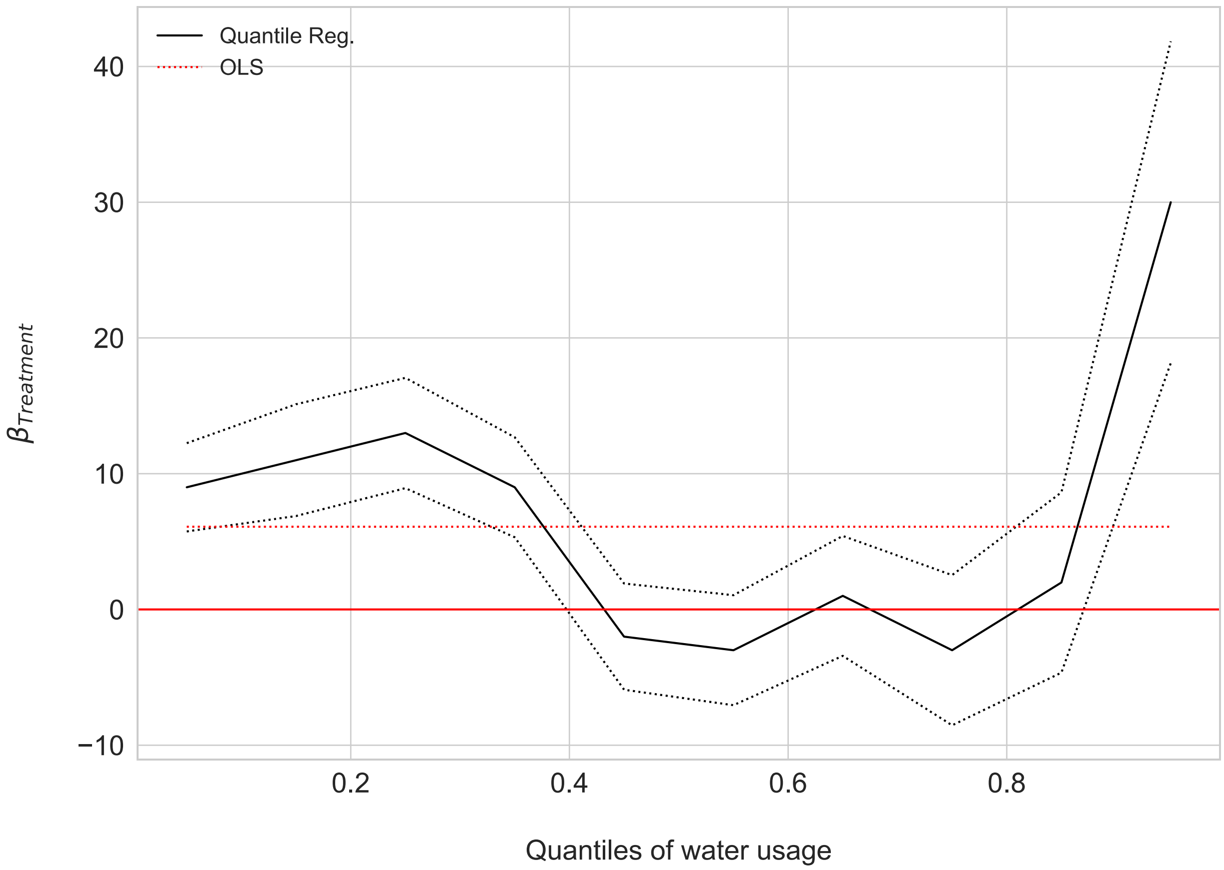

Figure 4 Quantile treatment effects for the treatment across the 10 months after the last home report was sent. The dotted red line represents the unconditional average treatment effect. It shows an increase in consumption because it has not controlled for baseline consumption in the same way the main model specification for the ATE has. The black line represents the unconditional quantile treatment effect. Black dotted lines represent the 95% confidence intervals. Standard errors were clustered on the household level.

4.6.3. Conditional Average Treatment Effects

Heterogeneity of treatment effects is also examined by interacting the treatment effect with decile bins of baseline water consumption. Therefore, when observing the treatment effect for the upper deciles of baseline consumption the focus is on households in the sample that populate the high end of the consumption distribution during the baseline period. Figure 5 below suggests that it is mostly households in the lower deciles that reduce their water consumption. These results are more interpretable than the quantile treatment effects because we can observe how the treatment effect differed based on the different household groups (based on baseline consumption).

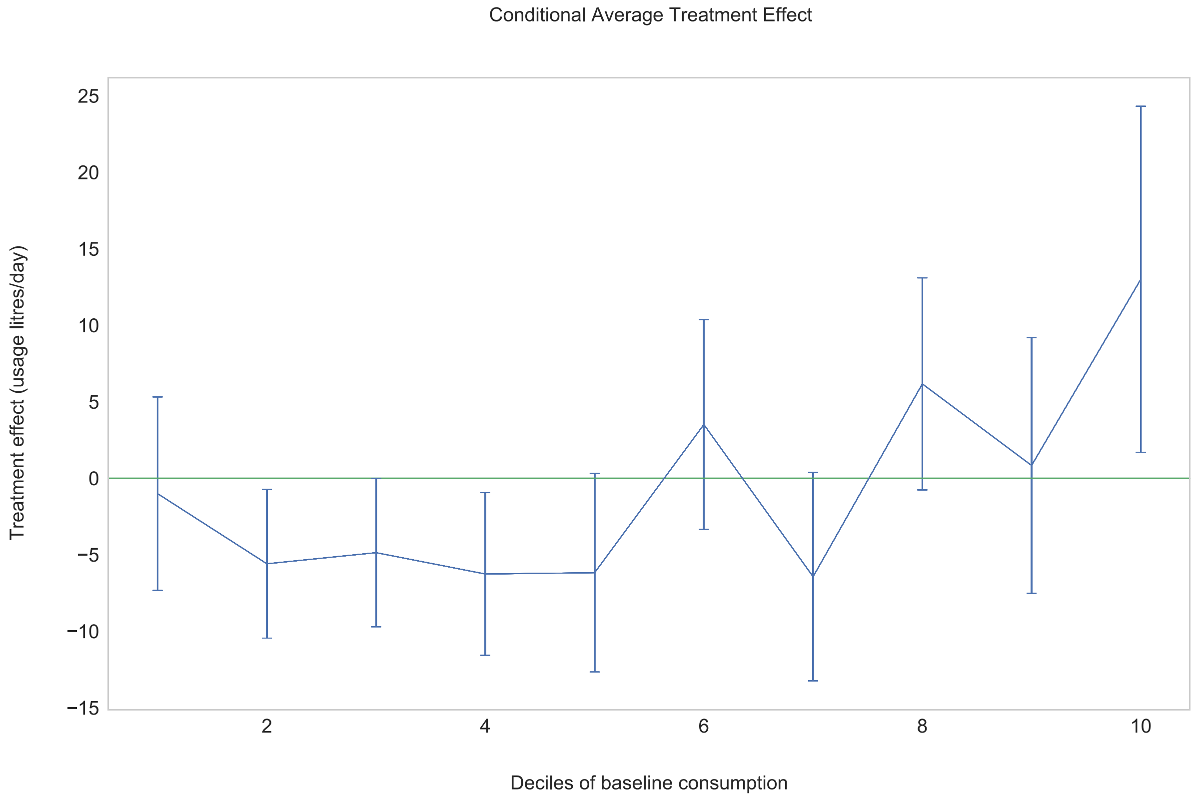

Figure 5 Graph showing the average treatment effect (y-axis) conditional on 10 bins of baseline consumption (x-axis) during the first 29 months of the programme. On the x-axis, 1 represents households with low water usage during the baseline period, while 10 represents the highest usage. The coefficients and standard error were extracted from a regression with interaction effects between treatment assignment and decile bins of baseline consumption.

4.6.4. Survey analysis

Following one year of the programme, South East Water recruited a survey company to interview a random sample of 200 customers in the treatment group. These surveys are conducted regularly as part of SEW's process of monitoring customer satisfaction. For this round of the survey, additional questions were included to enquire about the programme specifically. Questions for the survey can be found in the appendix. Out of the 200 customers surveyed, 87% recalled receiving the home water report, and around half of those read the report thoroughly. 60% of those who recalled receiving the report said that they were satisfied with the programme. When asked in an open question format which aspect of the report customers liked most, 90% mentioned the social norms aspect of the report. Customer account numbers were not collected in this survey; therefore, responses could not be linked to their consumption data.

4.7. Discussion

The current paper looks at the effects of a social norms based eco-feedback intervention on household water use in the UK. Results showed that, in line with the hypothesis, the social norms based eco-feedback interventions effectively reduce household water consumption by 1.8%. This effect appears to be highest during the summer months, which is likely because there is more scope for water conservation by reducing non-essential water use such as watering gardens during these months. The effect also appears to be stronger at the beginning of the programme than at the end. The 1.8% average treatment effect is smaller than those observed in the US and Colombia studies. This may be because water consumption in the UK is generally lower than that found in the US and Colombia, so there is less scope for improvement. This is similar to the findings in energy-focused interventions, that effect sizes in Europe are smaller than those in the US, due to the lower energy consumption in Europe (Andor et al., 2020; Bonan et al., 2020). The average water consumption of an individual living in the US is around 400 litres per day, whereas in the UK it is 141 litres per day.

Unlike that seen by Allcott and Rogers (Allcott & Rogers, 2014), these results do not show any 'action and backsliding' of treatment effects following the delivery of each home report. This could be due to the lower frequency in which the meters are read, which was six-monthly instead of daily. The lower frequency reads make it much harder to notice immediate and subtle changes in behaviour. Similarly, unlike that seen in Allcott and Rogers (Allcott & Rogers, 2014), no habit formation was observed following the end of the programme. Instead of ‘action and backsliding’ occurring on a monthly or quarterly level that gradually takes stock, the low frequency of six-monthly treatments may allow too much backsliding to occur so that habits are unable to be formed. This interpretation of a decay in treatment effect over the course of the programme should be made with caution because of the attrition in the number of households receiving the treatment over time and the general attrition of available data for households. The low frequency at which the treatment was delivered might also explain the smaller average treatment effect found in this experiment compared to other similar studies where treatment was delivered monthly (Bhanot, 2017, 2018; Brent et al., 2015; P. J. Ferraro et al., 2011; Jaime Torres & Carlsson, 2018). In Allcott (2011), treatment delivered at quarterly intervals had an effect size 0.5% smaller than when treatment was delivered at monthly intervals.

Assuming rank invariance, the analysis of quantile treatment effects suggests the presence of a boomerang effect. Households in the bottom 20% seem to increase their consumption quite sharply. These may be the households that have received feedback labelling them as the top 20% most ‘Efficient Neighbours’, which could licence them to increase their consumption. If this is the case, this would suggest that the intervention's injunctive norm component failed to counter the boomerang effect, counter to that found in other studies [15, 31]. Furthermore, those in the top 20% also increased their consumption sharply. It may be that certain households experienced psychological reactance to the social norms messaging, and chose to increase their consumption in response to feeling pushed to reduce their consumption by their utility (Brehm, 1989). These results differ from that found by Ferraro and Miranda (Allcott, 2011; P. J. Ferraro & Miranda, 2013), as neither interventions increased consumption across the entire distribution. The heterogeneity analysis based on CATE of baseline consumption suggests that households with lower consumption at baseline were decreasing their consumption the most. This would suggest that counter to the findings of the quantile treatment effects based on an assumption of rank invariance, a boomerang effect may not be present. The intervention is mostly only effective on households that were already relatively low consumers. This finding is in line with previous studies that also utilised injunctive norms to encourage already efficient consumers to keep reducing their consumption. Unlike those previous studies though, the current finding does not find any reduction in consumption for households that were high consumers at baseline. One explanation for these differences could be that households that already cared about conservation were affected by the motivational effects of the injunctive norm messaging, whereas high consumers were not fazed by the message of the descriptive norm or were experiencing psychological reactance. Alternatively, high consumers may have simply felt that the goal of reducing their consumption was too great and unachievable, leading to disengagement. Overall, as these results do differ from that of previous norms based eco-feedback interventions, it cannot be assumed that these interventions will bear the same results in all contexts. This is especially the case when studying the heterogeneity of the treatment effect. Different sub-populations will respond differently to these interventions, and these differences need to be better understood to fully maximise the impact of such treatment across all populations. At the very least, sub-populations that may negatively respond to these interventions may need to be identified and excluded from receiving the intervention.

4.8. Conclusion

This study demonstrates that while some aspects of the results may differ, social norms based eco-feedback is an effective intervention for reducing water household water consumption in the UK, even with infrequent 6-monthly communications. These effects may not, however, persist once the intervention is stopped. This study has several limitations, including the lack of additional measures of the households to help build a more precise model estimation and better interpret the findings and identify potential mechanisms. For example, not knowing when households are billed means that I cannot control for the effect of receiving a water utility bill on consumption. Not having more detailed information about the type of property or the occupants' demographics means that I am not able to control for the differences in their consumption patterns or attribute any heterogeneity in treatment effects to these group differences. A pre-intervention survey to gather this data was not conducted because it was not seen as commercially viable to Advizzo and SEW. The additional insights gained from such data would not be beneficial because this type of data would typically not be readily available. Therefore, when the intervention is scaled up, these additional insights would not be actionable. For example, even if it is found that pro-environmental people do not respond to these interventions, and so there is no need to target them, this characteristic would not be observable amongst households prior to an intervention when scaling the programme up.

An additional limitation, as previously mentioned, is that the low frequency of meter reads meant that changes in behaviour were more difficult to detect, and that the frequency of the treatment was also limited. This may be an argument for the promotion of smart meters that provide higher frequency reads. Finally, as discovered through the survey, not all households recalled receiving the treatment and only a portion of those that did mention they had read the home report thoroughly. If there were a way to better track which households had received and engaged with the home report, that data could be used to conduct instrumental variable analysis, which would give us a clearer estimate of the effect size.

Nonetheless, the current study provided additional evidence for policymakers to recognise the value of utilising social norms based eco-feedback to reduce household water consumption. The current study also helps to further generalise the potential effects of this type of intervention by demonstrating its effectiveness in a novel context.

Future studies may benefit from identifying ways to increase the impact and effectiveness of these interventions. This may be done through differences in how the intervention is delivered (e.g., greater frequency to increase its long term effects) or by leveraging the heterogeneity of the treatment effect to develop a more targeted approach to delivering the intervention.

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, *25*, 273–291. https://doi.org/10.1016/j.jenvp.2005.08.002

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*(10), 3003–3037. https://doi.org/10.1257/aer.104.10.3003

Andor, M. A., Gerster, A., Peters, J., & Schmidt, C. M. (2020). Social Norms and Energy Conservation Beyond the US. *Journal of Environmental Economics and Management*, *103*, 102351. https://doi.org/10.1016/j.jeem.2020.102351

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *Journal of Law, Economics, and Organization*, *29*(5), 992–1022. https://doi.org/10.1093/jleo/ews020

Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, *62*, 155–172. https://doi.org/10.1016/j.joep.2017.06.011

Bhanot, S. P. (2018). Isolating the effect of injunctive norms on conservation behavior: New evidence from a field experiment in California. *Organizational Behavior and Human Decision Processes*, *June*, 1–13. https://doi.org/10.1016/J.OBHDP.2018.11.002

Bicchieri, C., & Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*. https://doi.org/10.1007/s11127-019-00684-6

Bonan, J., Cattaneo, C., d'Adda, G., & Tavoni, M. (2020). The interaction of descriptive and injunctive social norms in promoting energy conservation. *Nature Energy*, *5*(11), 900–909. https://doi.org/10.1038/s41560-020-00719-z

Boretti, A., & Rosa, L. (2019). Reassessing the projections of the World Water Development Report. *Npj Clean Water*, *2*(1), 1–6. https://doi.org/10.1038/s41545-019-0039-9

Brehm, J. W. (1989). Psychological Reactance: Theory and Applications. *ACR North American Advances*, *NA-16*. https://www.acrwebsite.org/volumes/6883/volumes/v16/NA-16/full

Brent, D. A., Cook, J. H., & Olsen, S. (2015). Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials. *Journal of the Association of Environmental and Resource Economists*, *2*(4), 597–627. https://doi.org/10.1086/683427

Buchanan, K., & Russo, R. (2015). Going the extra green mile: When others' actions fall short of their responsibility. *Journal of Environmental Psychology*, *42*, 82–93. https://doi.org/10.1016/j.jenvp.2015.03.002

Carver, C. S., & Scheier, M. F. (2012). Cybernetic Control Processes and the Self-Regulation of Behavior. *The Oxford Handbook of Human Motivation*, *February 2017*, 1–30. https://doi.org/10.1093/oxfordhb/9780195399820.013.0003

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In *The handbook of social psychology, Vols. 1-2, 4th ed* (pp. 151–192). McGraw-Hill.

Czajkowski, M., Zagórska, K., & Hanley, N. (2019). Social norm nudging and preferences for household recycling. *Resource and Energy Economics*, *58*, 101110. https://doi.org/10.1016/j.reseneeco.2019.07.004

Ferraro, P. J., & Miranda, J. J. (2013). Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. *Resource and Energy Economics*, *35*(3), 356–379. https://doi.org/10.1016/j.reseneeco.2013.04.001

Ferraro, P. J., Miranda, J. J., & Price, M. K. (2011). The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment. *American Economic Review*, *101*(3), 318–322. https://doi.org/10.1257/aer.101.3.318

Ferraro, P., & Price, M. (2011). Using Non-Pecuniary Strategies to Influence Behavior: Evidence from a Large Scale Field Experiment. https://doi.org/10.3386/w17189

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*, *1*(1), 79–104. https://doi.org/10.1007/s12053-008-9009-7

Hussien, W. A., Memon, F. A., & Savic, D. A. (2016). Assessing and Modelling the Influence of Household Characteristics on Per Capita Water Consumption. *Water Resources Management*, *30*(9), 2931–2955. https://doi.org/10.1007/s11269-016-1314-x

Jaime Torres, M. M., & Carlsson, F. (2018). Direct and spillover effects of a social information campaign on residential water-savings. *Journal of Environmental Economics and Management*, *92*, 222–243. https://doi.org/10.1016/j.jeem.2018.08.005

Kandul, S., Lang, G., & Lanz, B. (2020). Social comparison and energy conservation in a collective action context: A field experiment. *Economics Letters*, *188*, 108947. https://doi.org/10.1016/j.econlet.2020.108947

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin*, *141*(6), 1205–1227. https://doi.org/10.1037/a0039650

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.137/0033-2909.119.2.254

Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging – A promising tool for sustainable consumption behaviour? *Journal of Cleaner Production*, *134*, 166–177. https://doi.org/10.1016/j.jclepro.2015.11.086

Mahler, H. I. M., Kulik, J. A., Butler, H. A., Gerrard, M., & Gibbons, F. X. (2008). Social norms information enhances the efficacy of an appearance-based sun protection intervention. *Social Science & Medicine*, *67*(2), 321–329. https://doi.org/10.1016/j.socscimed.2008.03.037

McKenzie, D., (2012). Beyond baseline and follow-up: The case for more T in experiments. Journal of Development Economics, 99(2), 210-221. https://doi.org/10.1016/j.jdeveco.2012.01.002

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative Social Influence is Underdetected. *Personality and Social Psychology Bulletin*, *34*(7), 913–923. https://doi.org/10.1177/0146167208316691

Perkins, H. W. (2002). Social Norms and the Prevention of Alcohol Misuse in Collegiate Contexts. *Journal of Studies on Alcohol/Supplement*, *14*(14), 164–172.
Reynaud, A., & Romano, G. (2018). Advances in the Economic Analysis of Residential Water Use: An Introduction. *Water*, *10*(9), 1162. https://doi.org/10.3390/w10091162
Rob Lawson, Marshallsay, D., DiFiore, D., Rogerson, S., Meeus, S., Sanders, J., & Horton, B. (2018, May 26). *The long term potential for deep reductions in household water demand*. https://www.ofwat.gov.uk/wp-content/uploads/2018/05/The-long-term-potential-for-deep-reductions-in-household-water-demand-report-by-Artesia-Consulting.pdf
Schultz, P. W., Messina, A., Tronu, G., Limas, E. F., Gupta, R., & Estrada, M. (2016). Personalized Normative Feedback and the Moderating Role of Personal Norms: A Field Experiment to Reduce Residential Water Consumption. *Environment and Behavior*, *48*(5), 686–710. https://doi.org/10.1177/0013916514553835
Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science*, *18*(5), 429. https://doi.org/10.1111/j.1467-9280.2007.01917.x
Sherif, M. (1936). *The psychology of social norms* (pp. xii, 210). Harper.
The 11 cities most likely to run out of drinking water—Like Cape Town. (2018, February 11). *BBC News*. https://www.bbc.com/news/world-42982959
Wada, Y., Flörke, M., Hanasaki, N., Eisner, S., Fischer, G., Tramberend, S., Satoh, Y., van Vliet, M. T. H., Yillia, P., Ringler, C., Burek, P., & Wiberg, D. (2016). Modeling global water use for the 21st century: The Water Futures and Solutions (WFaS) initiative and its approaches. *Geoscientific Model Development*, *9*(1), 175–222. https://doi.org/10.5194/gmd-9-175-2016

Appendix

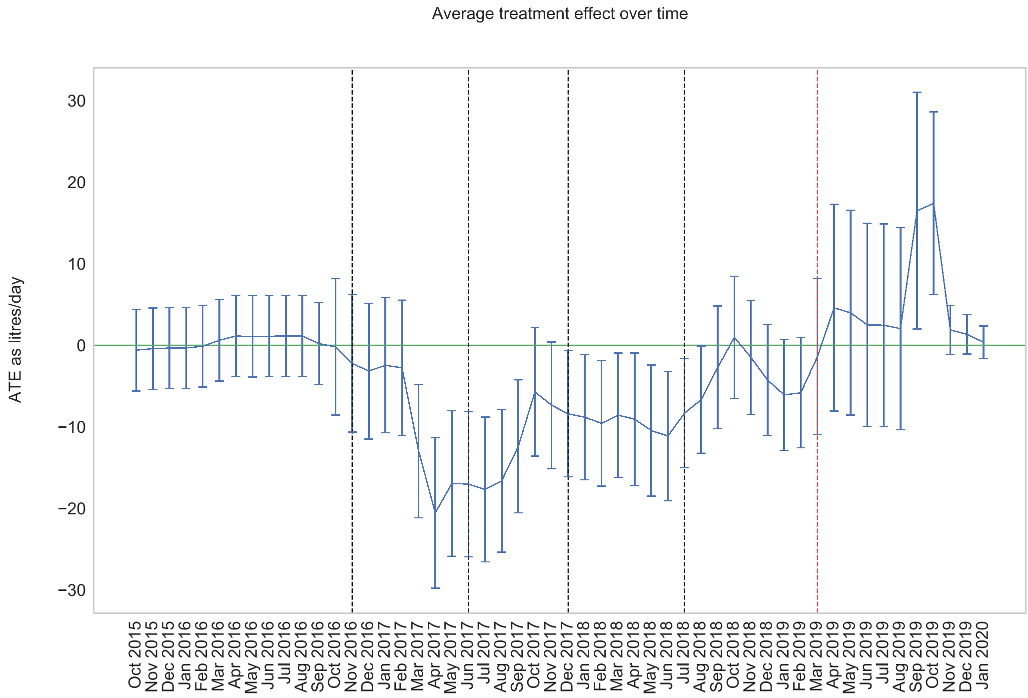

Figure A 1 Event study graph of changes in the effect size of the intervention across all months of the programme using data without the removal of outliers below and above the 1st and 99th percentile.

	Pre-treatment (Litres/day)	Treatment (Litres/day)	Post-Treatment (Litres/day)	No. of households (Treatment)	No. of households (Post-treatment)
Treatment	287.59 (142.58)	292.65 (165.75)	297.85 (163.15)	1716	1503
Control	279.05 (144.30)	291.30 (163.48)	291.75 (161.16)	1743	1517

Table A 1 Descriptive statistics of mean water consumption in litres per day across the baseline, treatment and post-treatment period, as well number of households in each period.

Q42 You were recently sent a 'water use' report from South East Water. This was one of **two** you have been sent in the last six months. The water use reports were a one-page printed reports sent by post, separate from your bill. TheyIt included a comparison of your water use with that of your neighbours'. The reports also included tips on how to save water.

Do you recall receiving a water use report?

1. Yes
2. No
3. I don't know/Can't remember
4. Did not receive any reports

Q43 **ASK IF Q42=1** Thinking of all the water use reports that you have received, what have you done with them? Did you... **READ CODES**

1. Read the reports thoroughly
2. Read some of the content
3. Glanced at the graphs
4. Did not look at the report at all
5. Don't know/Can't remember

Q44 **ASK IF Q43=4 or 5** Did someone else in the household read the reports?

1. Yes
2. No
3. Don't know/Can't remember

Q45 What do you think about South East Water sending information to customers in this way? **DO NOT READ – MULTICODE AS APPROPRIATE**

1. Good idea
2. Good to give information to customers
3. Informative
4. Helpful
5. Knew information already
6. Waste of time
7. Waste of money
8. Other, specify
9. Don't know/unsure **NOT WITH ANY OTHER CODE**

Q46 **ASK IF Q42= 1** On a scale of 1 to 5 where 1 is very dissatisfied and 5 is very satisfied, how satisfied are you with the water use programme? **IF REQUIRED:** The water use programme includes sending reports to customers, and giving them access to a web portal that helps them better understand and reduce their water consumption.

1. Very dissatisfied
2. Dissatisfied
3. Neither dissatisfied nor satisfied
4. Satisfied
5. Very satisfied
6. Don’t know/Not sure **DO NOT READ**

Q47 ASK IF Q42= 1 Thinking about the water use reports, please tell me whether you strongly disagree, Disagree, Neither Agree nor Disagree, Agree or Strongly Agree: **ROTATE**

		Strongly Disagree 1	**Disagree 2**	**Neither Agree nor Disagree 3**	**Agree 4**	**Strongly Agree 6**	**Don’t know 6**
a	The water efficiency tips in the water use reports are useful	1	2	3	4	5	6
b	The water use reports helps me make better decisions to use and save water	1	2	3	4	5	6
c	The water user reports will help me better monitor my water usage	1	2	3	4	5	6
d	The neighbour comparison in the water use reports helps me better understand my water usage	1	2	3	4	5	6

Q48 **ASK IF Q42=1** What aspects of the water use reports do you like the most? **OPEN: RECORD FULLY**

OPEN REPONSE

Don’t know/unsure

Table A 2 Subset of questions included in customer satisfaction survey to understand customer responses to the home water report

5. Paper 2: Optimal delivery of social norms based feedback for household water consumption

Abstract: Feedback interventions are an important approach to encourage reductions in household water and energy consumption but can vary in their efficacy. This paper investigates the different ways in which these interventions can be delivered through two field experiments on UK household water consumption. Overall, feedback interventions reduced consumption by around 2%. There were no significant differences in delivering the treatment monthly, quarterly, or bi-annually, but treatment was only effective when delivered by paper as opposed to by email. While almost all households reduced their consumption, treatment effects were heterogenous across different baseline consumption. Results suggest that interventions can be designed and targeted to maximise cost effectiveness.

5.1. Introduction

Providing feedback to consumers regarding their consumption of resources is a popular form of intervention to promote sustainable behaviours (Andor & Fels, 2018; Karlin et al., 2015; Sanguinetti et al., 2018). These feedback interventions are often combined with social information about the consumption patterns of their peers. These interventions have proven to be effective at reducing household energy consumption (Allcott, 2011; Andor et al., 2020; Ayres et al., 2013; Bonan et al., 2020; Costa & Kahn, 2013; Dolan & Metcalfe, 2013; Schultz et al., 2007), as well as household water consumption (Bhanot, 2017; Brent et al., 2015; Carlsson et al., 2020; P. J. Ferraro & Miranda, 2013; P. Ferraro & Price, 2011; Ramli, 2021; Schultz et al., 2016). These interventions have seen reductions in energy consumption by around 2% and around 4% in water consumption. The growing evidence base highlighting the efficacy and cost-effectiveness of these interventions has led municipalities and utilities worldwide to adopt them as a strategy for managing consumer demand. The proliferation of these interventions has highlighted the considerable heterogeneity in their treatment effects (Allcott & Mullainathan, 2010; Andor et al., 2020). This could be because the proliferation of these interventions means that there are operationally different ways to implement them based on the availability of technologies, for example, in areas where households have smart meters that record consumption at a higher frequency affords more granular information or more frequent communications. These differences in implementation have been found to have a moderating impact on the overall treatment effect (Karlin et al., 2015; Kluger & DeNisi, 1996; Sønderlund et al., 2014). The current study therefore explores how some of these differences in the way interventions are delivered can affect the overall treatment effect. This is done through two separate experiments in the context of water consumption in the UK. The first experiment looks at the difference in effects of different frequencies of treatment delivery, while the second experiment looks at the difference in effects of delivering the treatment through different mediums.

5.1.1. Intervention moderators

Looking at the wider literature of feedback interventions, several meta-analyses and reviews have found that the average treatment effects can vary quite widely, and have attributed this to the differences in the characteristics of these interventions. For example, Kluger and DeNisi (1996) developed the Feedback Intervention Theory by conducting a meta-analysis of 607 treatment effects of feedback interventions on a wide range of behaviours, and found that two-thirds of the interventions were effective at changing behaviour, while a third were ineffective. This meta-analysis was the first attempt at understanding how differences in the way feedback interventions are implemented can impact the treatment effects. More specific to the area of energy consumption, reviews by Fischer (2008), Karlin et al., (2015), and Sanguinetti et al., (2018) highlighted several components of feedback interventions that moderate its effects. These are the 1) the frequency at which the intervention is delivered, 2) the medium or format the information is delivered (e.g. letters, display units, emails), 3) the metric by which the intervention is framed (e.g. litres, price, or CO^2), 4) what the feedback is compared against (e.g. historical behaviour, social norms), 5) the granularity or level of detail of the feedback (e.g. the whole house, room, or by appliance), and 6) the duration by which the intervention is delivered.

This paper will focus on frequency and medium for two main reasons. First, these components may increase the treatment effect by increasing saliency to the treatment. That is, people may change their behaviour more or may attenuate their behaviour less, when they are exposed to the treatment more frequently. Additionally, people may engage with the treatment more if it is delivered in a medium that is either preferred or more impactful. Second, these components have significant operational costs associated with them, and so there may be diminishing returns of the treatment effect with increases in frequency or choice of medium. Sending paper mail includes printing and postal costs, and this is multiplied with higher frequency of feedback delivery. The other components have less of an impact on operational cost, or simply have technical limitations. Using social norms as a point of comparison has been extensively studied and has almost become the standard by which these feedback interventions are framed (Karlin et al., 2015; Nolan et al., 2008). For example, the California Public Utilities Commission restricted behaviour-based energy efficiency programmes to reports that included social norms based comparisons (Mahone & Haley, 2011). To avoid misalignment and confusion, the metric used to frame the feedback in the context of residential utility consumption is often determined by the metric used in the billing of the utility, and so there is less scope for changing this component. Furthermore, trying to deliver feedback that is highly granular is challenging due to technical limitations. Meters would need to be installed on the appliance level to be able to track consumption. An example of this is a feedback intervention on shower usage that utilised a meter directly installed on the shower itself (Tiefenbeck et al., 2019). While effective, the practicality of installing such a device on entire populations may not be feasible, whereas current feedback interventions can simply utilise house level meters already installed and used by utilities.

5.1.2. Medium: Email vs Paper

Since the first social norms based interventions for reducing residential utility consumption were delivered through door hangers, other mediums of delivery have been explored, including delivery through 'home report' delivered by paper mail to households (e.g. Allcott, 2011), home reports through emails (e.g. Dolan & Metcalfe, 2013), through an in-home display, a device that provides real time feedback of consumption on a digital screen (e.g. Schultz et al., 2015), or through a website or mobile application (e.g. Geelen et al., 2019). The main difference between email and paper mail compared to in-home displays and website, is that feedback information for the former two are 'pushed' on to people, while the latter two require people to 'pull' the information. The latter two mediums rely on its users to dictate the frequency at which they are treated by the intervention, and typically requires the availability of smart meters to be able to implement, which is not as widely available for water consumption. The focus here will therefore be on the use of paper mail or email. Furthermore, previous studies have mostly only looked at the use of paper mail. This is likely because a utility is assured to have the home addresses of customers, but less likely to have their email addresses, or may have not have in-home displays available.

The most important function of the medium of the intervention is its ability to engage the audience. If a household does not engage with the feedback information, it is unlikely they would benefit from its treatment effects. Mediums that are challenging to access or are too easily ignored would probably have lower effects on behaviours. Despite being quite a critical component of the success of these feedback interventions, only Dolan and Metcalfe (2013) have directly compared the two mediums of paper and email, and found that the intervention only worked when it was delivered by paper mail. Other reviews have found that computer based mediums were more effective, suggesting that it is because they are more interactive and are thus more engaging (Fischer, 2008; Karlin et al., 2015; Sanguinetti et al., 2018). The issue with this argument is that it assumes people a degree of willingness from people to log in to access the feedback voluntarily. A 2015 poll found that the average adult American has 500 unread emails (Moore, 2015). One factor to consider when delivering feedback interventions is the Ostrich Problem (Webb et al., 2013). People are not always motivated to monitor their goal progress (i.e., feedback), even if they rate that goal as something important. This can also be exacerbated if the individual believes that they are not performing well. With emails, it is often much easier to identify what the content will be through the email subject, and thus it is easier to ignore. This is especially the case if the feedback is not directly included in the body of the email and requires clicking through to a website. On the other hand, the content of letters is difficult to discern without opening it, therefore, an individual can be immediately treated by simply opening the letter to explore its contents.

5.1.3. Frequency

Since early research in feedback interventions, it has generally been assumed that the more frequent the feedback the more effective an intervention (Ilgen et al., 1979; Lam et al., 2011; Salmoni et al., 1984). More recent meta-analyses of feedback interventions studies suggest, however, that the moderating effect of frequency of feedback is unclear (Bond et al., 2012; Hysong, 2009; Karlin et al., 2015; Kluger & DeNisi, 1996). For example, while reviews by Fischer (2008) and Darby (2006) suggested that more frequent feedback leads to more behaviour change, meta-analyses by Karlin et al. (2015) and Kluger and DeNisi (1996) could not find a significant moderating effect of frequency of feedback on behaviour change. This has been suggested to be due to its interaction with other factors of the feedback intervention (Hysong, 2009; Karlin et al., 2015). In one experiment on household energy consumption run by Opower using paper home reports, Allcott (2011) found a non-statistically significant 0.4% difference in treatment effects between households that received home reports on a monthly basis compared to those that received reports on a quarterly basis. Looking at the wider pool of 17 experiments run by Opower that delivered the intervention at varying frequencies, the unweighted mean of the average treatment effects for monthly and bimonthly interventions were 2.2%, while the average for quarterly interventions were 1.7%. This suggests that the effect of frequency on average treatment effect may not be significant enough to justify the increased cost to print and post the reports. This then begs the question, how much less frequently can the intervention be delivered without compromising the average treatment effect? This is especially relevant in the context of water utilities in the UK where the majority of households only have meters that are read on a 6-monthly basis. The consideration of cost of delivering monthly feedback interventions goes beyond just printing and postage, but also includes the implementation of smart meters. Of course, the value of smart meters for utility companies does go beyond just being able to deliver feedback interventions at higher frequencies. An additional consideration about the role of frequency of feedback on treatment effects is that higher frequencies may lead to long lasting effects after the treatment has ceased. Allcott and Rogers (2014) found that higher frequency engagement showed less attenuation of treatment effect in between reports, and this led to a longer lasting habit formation. The authors argue that by delivering the treatment just around the time at which the treatment effect starts to attenuate, the saliency of the treatment essentially renews the treatment effect, and thus blocks the attenuation from occurring.

5.1.4. Social norms based feedback intervention and its cost

Social norms based feedback was first conceptualised by Schulz et al. (2007) when they found that leaving door hangers that provided personalised feedback information about a household's energy consumption combined with descriptive social norms was effective at encouraging households to reduce their energy consumption. From this initial study, the intervention has since been turned into a service product sold to utilities by various companies. One of the first of these, Opower, provides these intervention service at a cost of around USD 1-2 per household per year (Allcott 2011). The majority of this cost is likely to cover the price of printing and mailing the reports. This was estimated to have a cost-effectiveness ratio of about 3.3 US cents per kWh reduction in household energy consumption. WaterSmart, a similar company but with a focus on residential water consumption, charged utilities an average of USD 10 per household per year, with a cost-effectiveness ratio of USD 0.69 per cubic meter reduction in water consumption (Brent et al., 2015; Nauges & Whittington, 2019). These costs are typically based on the common approach of delivering feedback reports by paper mail on a monthly basis. As printing and postage makes up a majority of the cost, it is worth exploring how these differences in the way feedback interventions are implemented can affect their overall treatment effect, and to identify the method of implementation that provides the greatest return of investment for utility companies.

5.1.5. Aim

The current paper looks at two field experiments that test the differences in treatment effect when implementing a social norms feedback intervention with varying approaches of implementation. Experiment 1 looks at the difference in treatment effects when delivering the intervention on a monthly, quarterly or 6-monthly basis. As Allcott (2011) had only found non-significant differences between monthly and quarterly treatments, no difference is expected to be found between the monthly and quarterly treatment groups here, and by extension between all groups and the 6-monthly groups as well. Being able to confirm this, would mean a potentially large savings opportunity for utilities implementing these interventions. This study will be the first to test the impact of different frequencies on the effect of a feedback intervention in the UK context, and the area of water consumption. Experiment 2 looks at the difference in treatment effects when delivering the intervention via email and paper mail. As Dolan and Metcalfe (2013) found that letters were significantly more effective than emails, the same result is expected to be found here. This study will be the first to test this in the context of household water consumption. The two studies will also explore how these differences would affect the heterogeneity of the treatment effect.

5.2. Methodology

5.2.1. General experimental design

Two randomised field experiments were conducted in two regions of the UK through two different regional water utility companies, Anglian Water and South East Water. Water utilities in the UK are private companies that are run as regional monopolies, and as such, are heavily regulated by the water authority, OFWAT. The regulators have incentives in place to encourage utility companies to reduce overall water demand, especially in drought prone areas such as those where the two experiments are conducted. Houses are billed on a 6-monthly basis, with the average annual bill being around £415 a year. Average consumption in the UK is around 141 litres per person, per day. For comparison, the US has a daily consumption of around 400 litres per day, while Germany has 121 litres per day.

5.2.2. The feedback intervention

Households in the treatment condition received a 'home report' and were given access to an online portal that contained the same content of the home report. Both the reports and online portal were developed by Advizzo, a software-as-a-service company that works with utilities. The main feature of the home report was the social norms messaging in the form of a neighbour comparison graph. The neighbour comparison graph displays a comparison of the water consumption of the target household against the consumption of 'average homes, as well as the consumption of 'efficient homes'. The three groups a household can be placed in is therefore: 'more than average' – those consuming more than the mean of similar households, 'below the average' – those consuming less than the mean of similar households, and 'most efficient' – those consuming less than 80% of households. This comparison is displayed using a bar graph where the bars represent the amount of water consumed. Consumption is displayed as cubic litres (m^3) so that it is aligned with what households see in their bill payment. This neighbour comparison graph serves the purposes of delivering descriptive social norms and has been used widely in previous research (e.g., Schulz et al., 2007; Alcott 2009; Ayers et al., 2013). In addition to this, an injunctive norm is also displayed in the form of a series of 'smiley faces' with three labelled levels, 'more than average', 'Good' and 'Great' to counter any possible boomerang or coasting effect (Schulz et al., 2007). Finally, the home report also included three tips of ways to reduce water consumption. Figure 6 provides an example of the report. The home reports were sent to households on a monthly, 3-monthly or 6-monthly basis, and through paper mail or email, all depending on the experimental group assignment. The control groups in both experiments did not receive any communications.

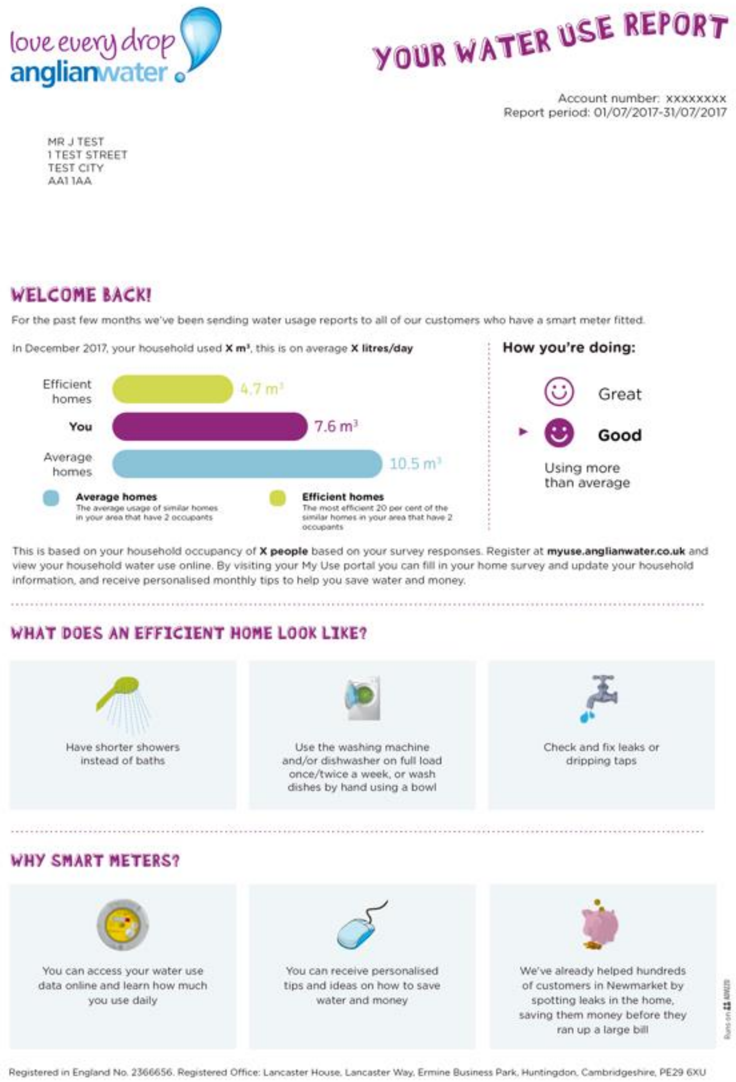

Figure 6 Example of home report sent to households

5.2.3. General sample selection

Households had to meet a set of criteria to be included in the programmes. These included, 1) having a working water meter, and having sufficient number of meter reads over their respective interval period (1/3/6 months) prior to selection to be able to generate the social norms, 2) the meter read was within an upper bound of 10,000 litres a day (i.e. the household has a leak or a meter read error) as determined by the respective water utility, 3) the customer is not on a social tariff, 4) the customer has not opted out of communications, 5) the customer account is residential, and 6) the customer account has not been closed. Before each treatment interval, a home report is only generated for households that are not excluded by this criteria. Even if a household is excluded from receiving a home report, their data continue to be collected as long they had not closed their account or had a meter read error.

5.2.4. General treatment

The dependent variable used is water consumption in litres per day. This was calculated by taking the difference between two meter reads to work out consumption over that period (monthly for the frequency experiment and 6 monthly for the medium experiment), then dividing it by the number of months within that period. This was then subsequently divided by the number of days within each month. Baseline consumption for each treatment month was determined by looking at consumption amount during the corresponding month over the baseline period.

There were three further exclusion criteria prior to data analysis. First, households that did not have any data over the baseline period were removed. Second, households that did not have any data over the treatment period were removed. Finally, outliers were removed at the 99 percentile. These were deemed as outliers because the level of consumption was unrealistically high for a household to consume in a day. For example, the frequency experiment had households with consumption levels at 2,500 litres per day, and the medium experiment had households with 80,000 litres per day. Both are significantly higher than the average UK household consumption of 349 litres per day.

5.2.4.1. Frequency experiment

Participants for experiment 1, testing the different frequencies of feedback delivery, comprised of customers from Anglian Water, a water utility operating in the East of England.
An initial sample of 22,000 households from the Colchester region of the UK were selected to be included in the study, but only 13,047 were qualified, based on the selection criteria, to receive the treatment. These were then equally randomised on the household level into a control and three treatment groups in May 2017. Consumption data from household water meters were collected on a monthly basis through the use of a refuse truck with a remote meter reader attached, driving around the region. Only 11 months of baseline consumption data was available from July 2016 until the start of treatment in June 2017. Baseline consumption for June 2016 was therefore imputed using consumption data from the next available month. Following the delivery of the first home report, additional reports were subsequently sent out every month, 3-

months or 6-months depending on the treatment group the households were assigned to. After one year of the intervention, only households in the monthly treatment group that had shared their email address with Anglian Water continued to receive the treatment. Households that had not shared their email address, as well as households in the 6-monthly and 3-monthly group, stopped receiving the treatment. Despite no longer receiving the treatment, consumption data for these households continued to be collected. Appendix Figure 1 shows the number of home reports sent across the entire programme period.

Out of the sample of 13,047 households, 425 were removed for not having data available in the baseline period, and 26 households were removed for not having any data in treatment period. 4,901 rows of data beyond the 99th percentile were treated as outliers and removed from the dataset. The final dataset consisted of 12,586 households. Two regression analyses on attrition using number of available months as well as a binary indicator of attrition on treatment assignment were conducted and showed no statistically significant differences (Appendix Table 3). Furthermore, baseline consumption between the treatment and control groups were not statistically significantly different and were therefore balanced at randomisation. Balanced baseline consumption can also be observed in the event study graph of Figure 7. Descriptive statistics are presented in Table 3.

5.2.4.2. Medium experiment

Participants for experiment 2, testing the different mediums of feedback delivery, comprised of customers from South East Water, a water utility operating in the South East region of England. Consumption data were manually collected from households on a 6-monthly basis. Twelve months of consumption data between March 2016 and March 2017 was used as the baseline period. Home reports were sent around every 8 months starting from March 2017, with a final report being sent in March 2019. Consumption data up until September 2019 was made available for analysis.

An initial sample of 20,000 households that had shared their email addresses with the utility were selected to be included in the study. Following the exclusion of households based on the criteria discussed above, 18,896 households were randomly assigned into either a treatment group where households received the home report by email, a treatment group where households received the home report by paper, or a control group that received no communication. Half the sample were assigned to the control group while the other half were split between the two treatment groups. 545 households that had no consumption data in the baseline period were removed, and a further 37 households that had no consumption data in the treatment period were also removed. Finally, 7,130 rows of consumption data beyond the 99^{th} percentile were treated as outliers and were removed. The final data set therefore consisted of 18,314 households. There were no significant differences in attrition between the three groups (Table A3). Descriptive statistics are presented in Table 3.

Group/experiment	Pre-treatment (Litres per day)	Post-treatment (Litres per day)	Number of households	Number of households (Year 1)	Number of households (subsequent years)
Frequency					
Monthly	297.28 (222.49)	277.40 (175.97)	3333	3311	3137
3-monthly	297.24 (263.06)	277.74 (175.91)	3255	3237	3088
6-monthly	297.29 (208.18)	281.59 (173.31)	2847	2845	2743
Control	290.38 (215.50)	281.43 (176.32)	3141	3122	2994
Medium					
Email	290.23 (155.87)	296.23 (146.94)	4565	4563	4183
Paper	289.00 (181.50)	289.11 (147.44)	4618	4612	4226
Control	287.82 (150.30)	293.48 (147.42)	9120	9108	8288

Table 3 Descriptive statistics for both experiments showing the average consumption of each group in litres per day, as well as the number of households in each group. Balance checks available in Appendix Table 3.

5.3. Results

Average treatment effects

To estimate the average treatment effects for both experiments, I use a regression specification similar to that of Allcott and Rogers (2014).

$$LPD_{it} = T_i + LPDpre_{it} + \pi_m + \epsilon \quad (5)$$

Where LPD_{it} is household $i's$ water consumption in litres per day on month t. T_i is the treatment indicator. $LPDpre_{it}$ is average water consumption in the matching calendar month in the baseline period, and π_m are month and year fixed effects. All analyses will include the clustering of standard errors on the household level.

5.3.1. Frequency experiment

As the treatment was delivered to the majority of the sample for the first 12 months only, the analysis is split to first focus on this initial period, the subsequent 23 months, and then all the months together. Results presented in Table 3 shows that only households in the monthly and 3-monthly treatment groups statistically significantly reduced their water consumption over the first 12 months period when the reports were being sent. Households in both these treatment groups reduced their consumption by around 6 litres per day, or a 2% reduction compared to the control group. The subsequent 23 months showed an increase in effect size with households in the monthly group reducing their consumption by 8 litres/day, and those in the three monthly group reducing their consumption by 7 litres/day. A pairwise t-test for equality of coefficients shows that there were no differences between these three treatment groups. As can be seen in Figure 7, following a drop in consumption during the first year while reports are being delivered, the treatment effect seems to attenuate after reaching its peak during the end of the first year of treatment.

	Treatment period (12 months) litres/day	Post-treatment period (23 months) litres/day	All months (35) Litres/day
1 monthly	-6.329*** (2.083)	-8.103*** (2.535)	-7.415*** (2.226)
3 monthly	-6.613*** (2.135)	-7.285*** (2.575)	-7.024*** (2.60)
6 monthly	-1.711 (2.101)	-4.349* (2.512)	-3.237*** (2.228)
3 monthly =1 monthly	-2.85 (2.154)	0.818 (2.591)	0.391 (2.286)
6 monthly = 1 monthly	4.617 (2.054)	3.754 (2.468)	4.178 (2.189)
6 monthly = 3 monthly	-4.902 (2.170)	2.936 (2.566)	-3.788 (2.280)
Number of households	12,515	11,962	12,576
Observations	145,299	217,717	363,016
R^2	0.474	0.366	0.406
F statistic	69.01	50.25	55.99

Table 4 Results of ATE in litres per day for the frequency experiment. Standard errors in parentheses. The dependent variable is household water consumption in litres per day. Results based on model (1) controlling for baseline consumption and time. Rows 4-6 in table are the results of pairwise t-test for equality of coefficients to identify whether each treatment group were statistically significantly different from each other. *p<0.1, **p<0.05, ***p<0.001

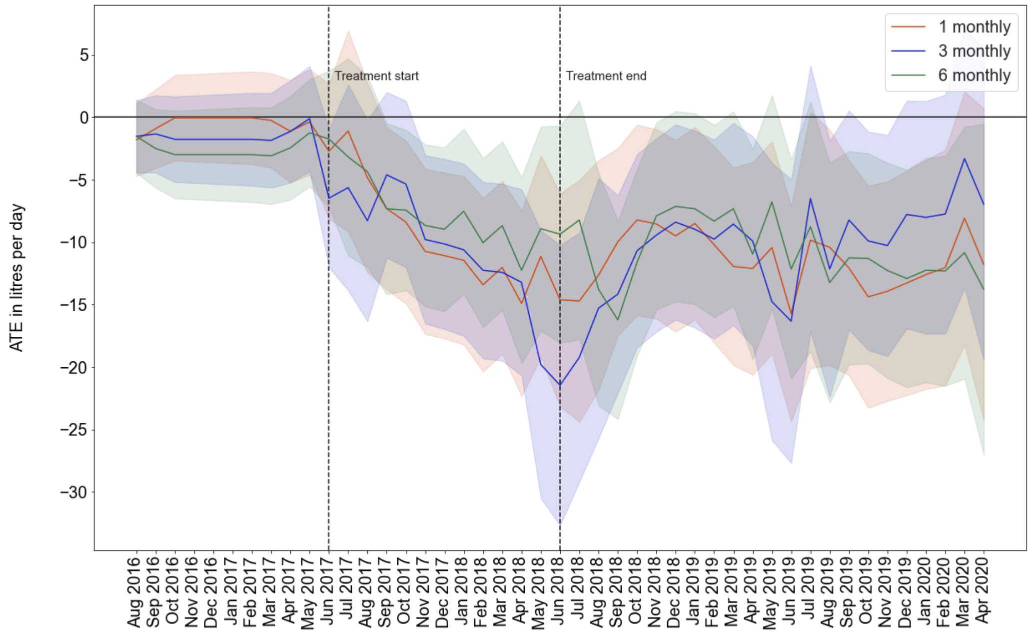

Figure 7 Event study graph of the average treatment effect in litres per day across all months before and after the intervention for all three treatment groups in the frequency experiment. The coefficients and confidence intervals were generated from interacting treatment assignment with monthly dummy variables while controlling for baseline consumption.

5.3.2. Medium experiment

Results in Table 5 highlights that only households that received the treatment by paper mail showed a statistically significant reduction in consumption compared to the control group. The difference in consumption between the two treatment groups, email and paper, were statistically significantly different. Households in the paper group reduced their consumption by around 5 litres a day, equivalent to a 1.75% reduction. As can be observed in Figure 8 of an event study graph of the treatment effects over time, households that received the reports by paper mail saw an initial reduction in consumption that persisted over time, while those that received the reports by email never really reduced consumption by a significant amount.

	31 months
Intercept	108.444 (16.81)
Email	1.204 (1.201)

Paper	-5.142*** (1.209)
Baseline consumption	0.642 (0.059)
Email = Paper	-6.346*** (1.429)
Number of households	18,303
Observations	489,977
R^2	0.488
F statistic	28.18

Table 5 Results of ATE analysis for the medium experiment. Standard errors in parentheses. The dependent variable is household water consumption in litres per day. Results based on model (1) controlling for baseline consumption and time. *p<0.1, **p<0.05, ***p<0.001

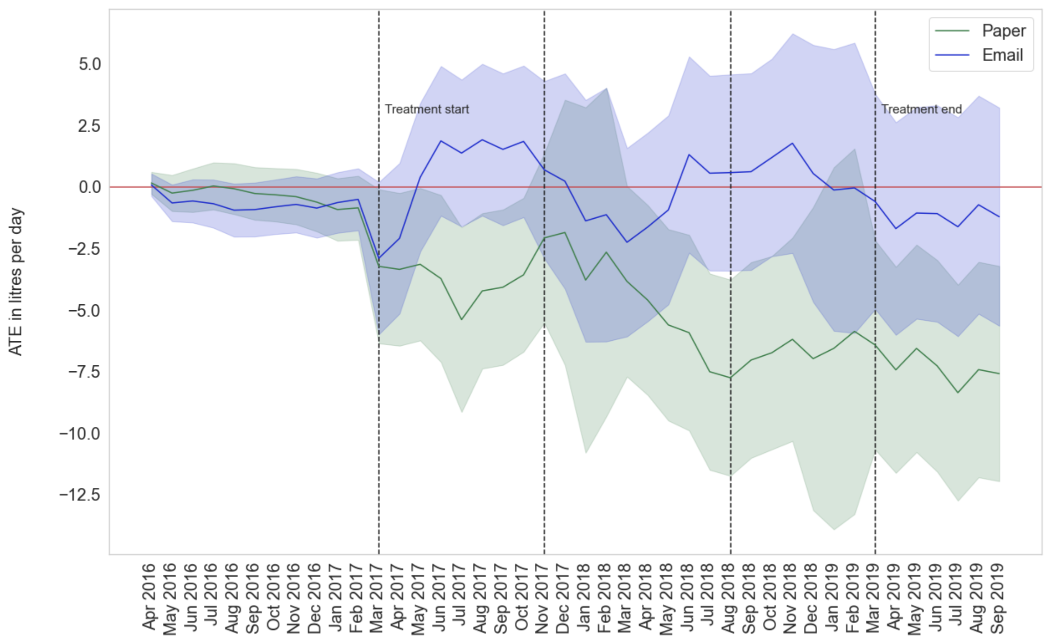

Figure 8 Event study graph of both the pre and post treatment period for both treatment groups of the medium experiment. Coefficients and confidence intervals were generated by interacting treatment assignment with monthly dummy variables while controlling for baseline consumption.

5.3.3. Heterogeneity

Beyond identifying the optimal method of delivering feedback interventions, studying the heterogeneity of treatment effects can further improve the targeting of the intervention to maximise the effects of the treatment. To do so I employ two methods of analysis, quantile regressions and conditional average treatment effect. Quantile regressions are useful for studying heterogeneity because rather than just estimating the treatment effect on the mean of the outcome of interest, quantile regressions allow the estimation of the effect on the median of this outcome alongside the full range quantiles, thus allowing the study of how the treatment effect may vary across the different aspects of the distribution of the outcome of interest. Here I use an unconditional quantile regression, $LPD_{it} = T_i$, for each treatment group separately (Appendix Equation 6). One limitation of quantile regressions is that due to rank invariance, it is difficult to identify specific groups within its output. Furthermore, heterogeneity can also be studied by looking at the conditional average treatment effect (CATE) of baseline consumption. This is often complementary to quantile regressions which are unable to make statements about groups of the sample due to rank invariance. CATE on the other hand are more interpretable to enable statements to be made about how different levels of baseline consumption are affected by the treatment. To run CATE, I use the same model specification as that for the ATE, with the addition of an interaction term between treatment group assignment and deciles of baseline consumption (Appendix Equation 7).

5.3.3.1. Experiment 1: Frequency

Figure 9 of the quantile treatment effects for both the monthly and 3 monthly groups shows that the entire distribution of consumption is not shifting downwards, and so there is no aggregate reduction in consumption across all households. Rather, the middle of the distribution seems to be attenuating the treatment effect, with a small amount of mass in the 3 monthly group moving above zero, suggesting an increase in consumption. The bottom tail is moving downwards but not by much, which is expected given that there is a lower bound to consumption. The biggest reduction is seen in the upper tail, suggesting that those with the highest consumption are reducing the most. This can be observed in the Figure 10 of CATE for the 3 monthly group, but less so for the monthly group where it seems that greatest reduction in consumption is seen by those in the 5th and 7th deciles. The 6 monthly group was omitted here to focus on monthly and

3 monthly, because results of the 6 monthly group were not statistically significant in the first year period.

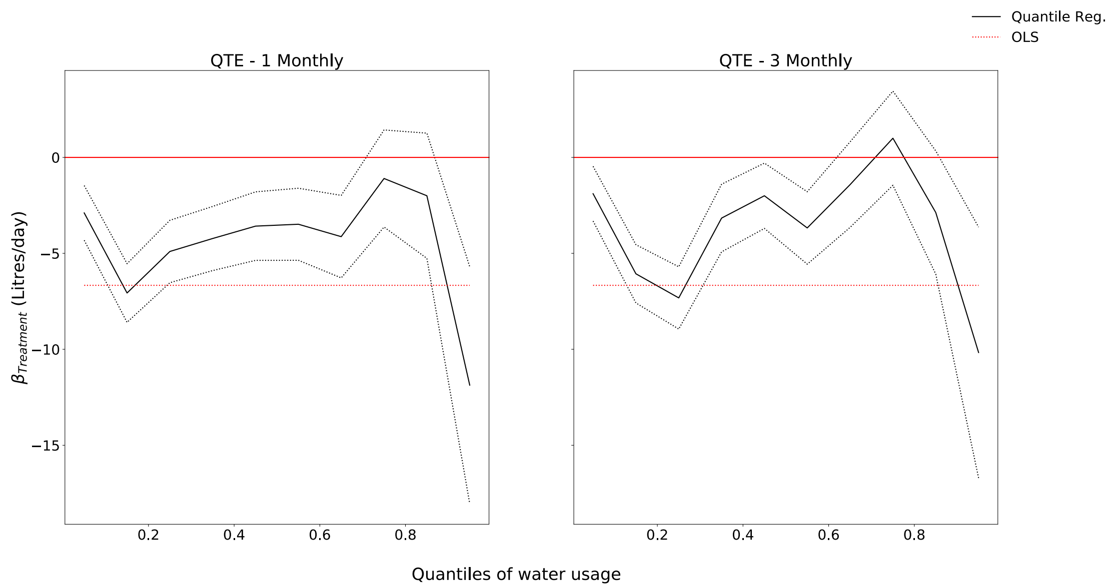

Figure 9 Quantile treatment effects of the monthly and 3 monthly groups.

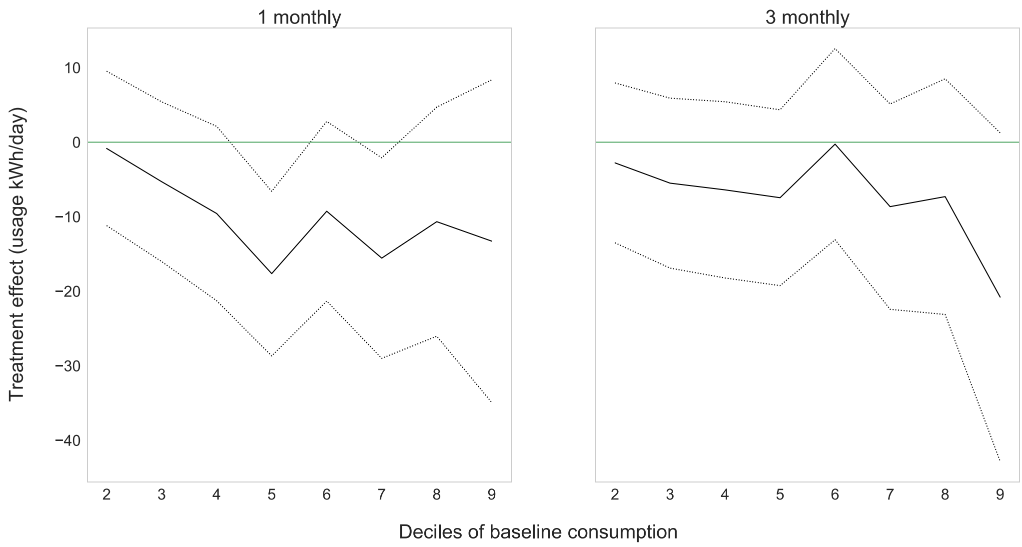

Figure 10 CATE of the monthly and 3-monthly group

5.3.3.2. Experiment 2: Medium

Looking at the results for the quantile regression analysis for the medium experiment in Figure 11, as expected, the bottom percentiles show minimal reductions in consumption, limited by the lower bounds of consumption. The middle distribution shows an S shape, but should not over interpreted as it could just be noise. Assuming this part of the distribution is flat, there is an expected uniform shift downwards by around 6 litres per day. The top of the distribution however is demonstrating an attenuation of the treatment effect. This is somewhat concerning as the top of the distribution typically has the most opportunity to reduce by more. It could be that high consumers are not reducing as much. This is further demonstrated by looking at the CATE in Figure 12 which shows that while the top and bottom deciles of baseline consumption show no significant reduction in consumption, the greatest reductions come from households in the middle deciles with those in the 6th decile reducing by around 16 litres per day.

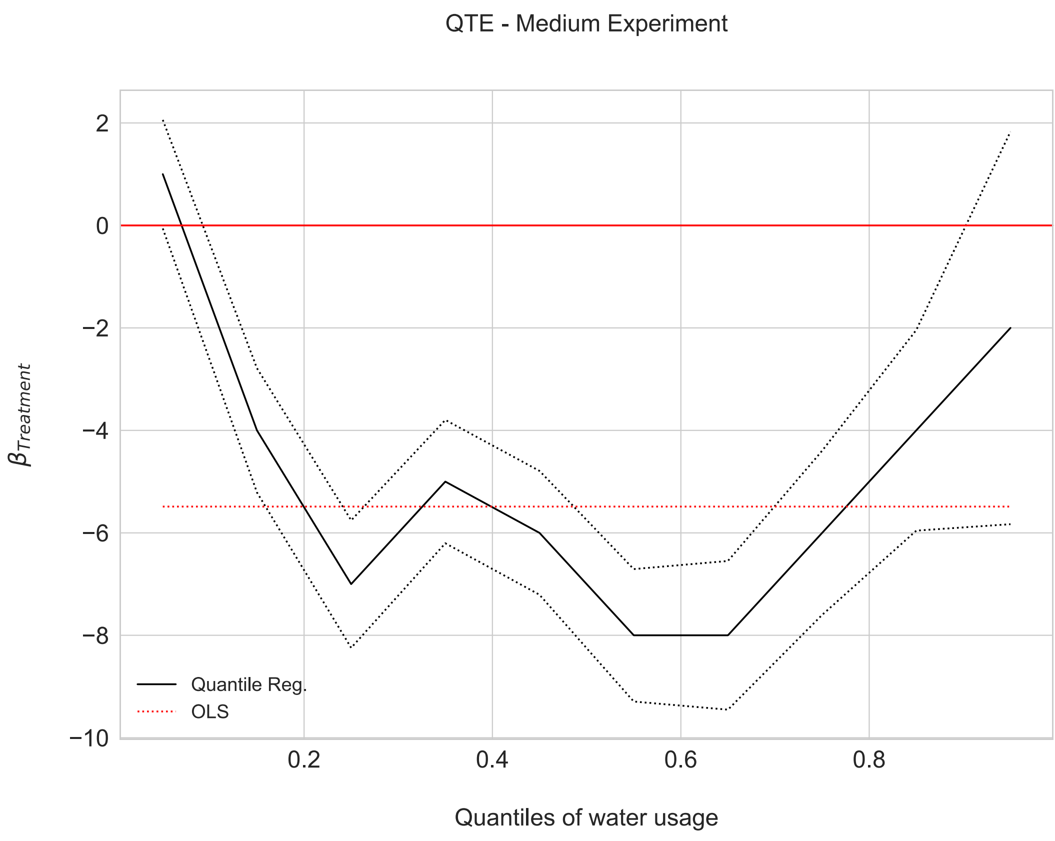

Figure 11 Quantile treatment effects of the paper group

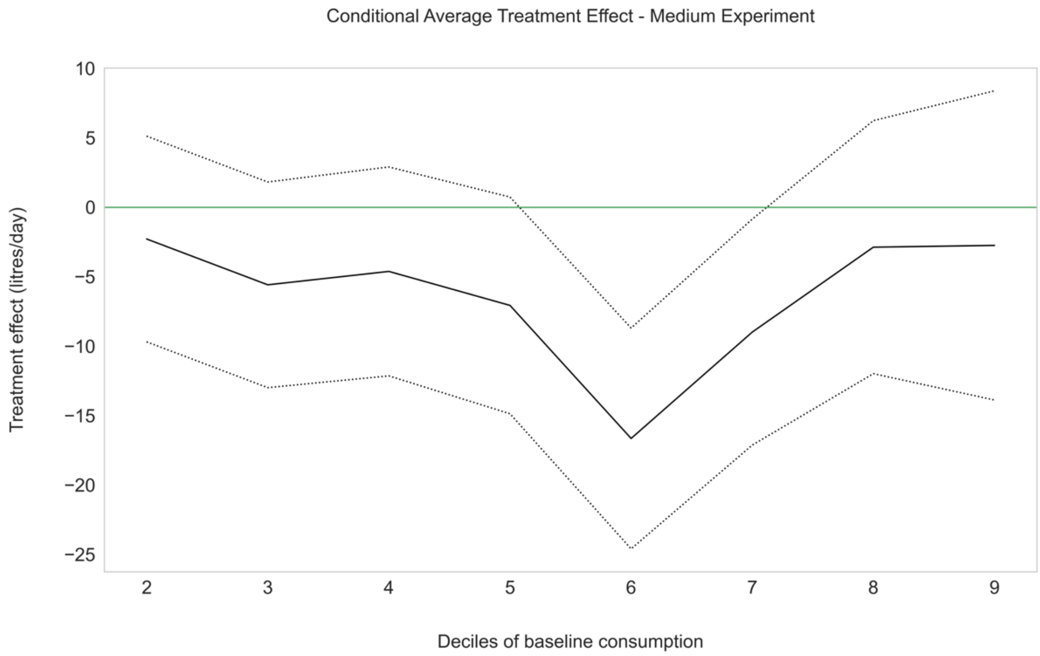

Figure 12 CATE of the paper mail group

5.3.4. Cost effectiveness

The mean water savings of combining the average treatment of a 3 monthly treatment delivered by paper mail from both experiments for one year of treatment is around 7 litres/day or 2,555 litres/year. The cost of generating, printing and posting a report is £0.80/household per report, and £3.20 for four reports a year. This means that utility companies can achieve a savings 798.43 litres for every pound spent per household. For households, based on Anglian Water's water pricing for the year 2021-22 of £1.6015/cubic metre, an intervention like this would reduce a household water bill by around £4. For a household, these savings are probably not significant, however, in aggregate, these reduction in consumption has significant value to utility companies. According to Nauges and Whittington (2019), the cost savings achievable by a water utility from a feedback intervention such as this comes from not requiring the electricity and chemicals used to deliver water. At most, this is around 25% of operation and maintenance, or 10% of the total average cost of producing a single cubic metre of water.

Cost effectiveness can also be increased by targeting households based on heterogeneity. For example, households in the 7th decile of baseline consumption in the Frequency experiment reduced consumption by 11 litres/day, which is equivalent to 1,255 litres of savings for ever pound spent. If the utility were to only target this segment of 943 households, they could achieve an annual saving of 3,786,145 litres of water at the price of 3017.60 for delivering the treatment, resulting in 1,254.7 litres saved per pound spent, versus the 798.44 when delivered to all households.

5.4. Discussion

The current paper explores the moderating effects of different components of social norms based feedback interventions on household water consumption in the UK. Two experiments were conducted to determine whether the frequency or the medium at which the intervention is delivered would affect the impact of the intervention. Results from both experiments first showed that the intervention is effective at reducing household water consumption in the UK. The frequency experiment reduced water consumption by around 2%, while the medium experiment reduced water consumption by 1.75%. These are similar to the 1.8% saving found by Ramli (2021) in the UK water context. These are relatively smaller effects when compared to the 5% savings achieved in the US (Ferraro and Price, 2013; Brent et al., 2015).

Quantile regressions also indicated that while there were heterogeneity in the treatment effect, the treatment was successful at encouraging almost everyone to reduce their water consumption in both experiments. Even if there are some that increased their consumption, it is balanced out by others that reduced their consumption by significant amounts. This suggests that there was no rebound effect where low consumers licensed themselves to increase their consumption. This is confirmed by the results of the CATE of baseline consumption. Furthermore, unlike many studies in both water and energy, the current experiments do not find that the most savings were achieved by those with the highest baseline consumption, but instead were achieved by those in the middle to upper deciles.

Results from the frequency experiment was in line with the findings of Allcott (2011) in that delivering the treatment at monthly or 3-monthly frequencies resulted in very similar treatment effects, and that the two groups were not statistically significantly different from each other. These treatment effects also cannot be distinguishable statistically from that of the 6-monthly group, despite its much smaller coefficient. This is unsurprising as the Medium experiment was able to achieve water savings despite treatment being delivered on a 6-monthly basis. Looking at Figure 8 of the event study graph for the Medium experiment, there is a gradual decrease in water savings (or an increase in treatment effect) over a long period. It may be the case that had the frequency experiment been extended beyond one year, the 6-monthly group may have achieved more comparable treatment effects. One argument for higher frequency of treatment is that found in by Allcott and Rogers (2014), where the treatment effect is strongest at the point at which households receive the reports, but would then decay soon after, up until they had received the subsequent report. I am unable to formally confirm this here as daily meter reads were not available, but it seems like the treatment effect for all three treatment groups actually increases, on average, in the 23 months after households stop receiving reports. Looking at the event study graph for the frequency experiment, it is unclear whether the treatment effect would attenuate or simply plateau after those 23 months. Therefore, while there is evidence to suggest that frequency does not have an effect on treatment, there is not enough evidence here to determine whether monthly reports are more likely to have lasting effects beyond the treatment.

Results from the medium experiment confirmed the findings of Dolan and Metcalfe (2013) by showing that delivering feedback interventions through email was not effective. This may simply be because emails are much easier to ignore compared to letters since one can often assume the content of an email just by looking at the sender's name or the subject line. Whereas a letter, if sent in a plain envelope, is more ambiguous until it is opened. Furthermore, the email treatment required people to click through to a website to receive the treatment, and so it might be the case that this additional layer of friction may have discouraged people from engaging with the feedback.

Together, these results suggest that within these contexts, the optimal design of social norms based feedback is to deliver paper mail based reports on a 3 to 6-monthly basis, over a long period, while mostly targeting households based on their baseline consumption. It might be the case that households with alternative baseline consumption may benefit from alternative interventions. There are two main limitations to the current study. Firstly, the lack of additional variables about the households means that I am unable to conduct sub-group analyses to identify additional heterogeneities in the treatment effects. It may be that certain portions of the population may respond to email over paper, which would help to reduce cost while increasing effectiveness. Second, the period after the last paper report was sent for the frequency experiment had some bias in that households in the monthly group that had shared their email address continued to receive the treatment via email. These numbers were very low, and still did not make the monthly group statistically different from the other two treatment groups. Future studies could therefore explore alternative ways to make emails more engaging as this would help to further reduce the cost of these interventions. Alternatively, future studies could also explore other mediums of engaging with households such as apps or text messages. The current study only compares email to paper, and so the conclusion found here are limited to that set of choices. It is therefore important to explore new ways of delivering this treatment in a way that is highly salient and engaging, while being cost effective.

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003

Allcott, H., & Mullainathan, S. (2010). Behavioral Science and Energy Policy. *Science*, *327*(6970), 1204–1205. https://doi.org/10.1126/science.1180775

Allcott, H., & Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, *104*(10), 3003–3037. https://doi.org/10.1257/aer.104.10.3003

Andor, M. A., & Fels, K. M. (2018). Behavioral Economics and Energy Conservation – A Systematic Review of Non-price Interventions and Their Causal Effects. *Ecological Economics*, *148*, 178–210. https://doi.org/10.1016/J.ECOLECON.2018.01.018

Andor, M. A., Gerster, A., Peters, J., & Schmidt, C. M. (2020). Social Norms and Energy Conservation Beyond the US. *Journal of Environmental Economics and Management*, *103*, 102351. https://doi.org/10.1016/j.jeem.2020.102351

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *Journal of Law, Economics, and Organization*, *29*(5), 992–1022. https://doi.org/10.1093/jleo/ews020

Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, *62*, 155–172. https://doi.org/10.1016/j.joep.2017.06.011

Bonan, J., Cattaneo, C., d'Adda, G., & Tavoni, M. (2020). The interaction of descriptive and injunctive social norms in promoting energy conservation. *Nature Energy*, *5*(11), 900–909. https://doi.org/10.1038/s41560-020-00719-z

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. https://doi.org/10.1038/nature11421

Brent, D. A., Cook, J. H., & Olsen, S. (2015). Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials. *Journal of the Association of Environmental and Resource Economists*, *2*(4), 597–627. https://doi.org/10.1086/683427

Carlsson, F., Jaime, M., & Villegas, C. (2020). Behavioral spillover effects from a social information campaign. *Journal of Environmental Economics and Management*, 102325. https://doi.org/10.1016/j.jeem.2020.102325

Costa, D. L., & Kahn, M. E. (2013). Energy conservation 'nudges' and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, *11*(3), 680–702. https://doi.org/10.1111/jeea.12011

Darby, S. (2006). *The effectiveness of feedback on energy consumption* (p. 24). Environmental Change Institute. https://www.eci.ox.ac.uk/research/energy/downloads/smart-metering-report.pdf

Dolan, P., & Metcalfe, R. (2013). Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation. *Centre for Economic Peformance Discussion Papers*, *1222*. https://doi.org/10.2139/ssrn.2589269

Ferraro, P. J., & Miranda, J. J. (2013). Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. *Resource and Energy Economics*, *35*(3), 356–379. https://doi.org/10.1016/j.reseneeco.2013.04.001

Ferraro, P., & Price, M. (2011). Using Non-Pecuniary Strategies to Influence Behavior: Evidence from a Large Scale Field Experiment. https://doi.org/10.3386/w17189

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency*, *1*(1), 79–104. https://doi.org/10.1007/s12053-008-9009-7

Geelen, D., Mugge, R., Silvester, S., & Bulters, A. (2019). The use of apps to promote energy saving: A study of smart meter–related feedback in the Netherlands. *Energy Efficiency*, *12*(6), 1635–1660. https://doi.org/10.1007/s12053-019-09777-z

Hysong, S. (2009). Meta-analysis: Audit and feedback features impact effectiveness on care quality. *Medical Care*, *47*(3), 356–363.

Ilgen, D., Fisher, C., & Susan, M. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*, 349–371. https://doi.org/10.1037/0021-9010.64.4.349

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin*, *141*(6), 1205–1227. https://doi.org/10.1037/a0039650

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the 'more is better' assumption. *Organizational Behavior and Human Decision Processes*, *116*(2), 217–228. https://doi.org/10.1016/j.obhdp.2011.05.002

Mahone, A., & Haley, B. (2011). *Overview of Residential Energy Feedback and Behavior-based Energy Efficiency* (No. 1219691; p. 1219691). https://doi.org/10.2172/1219691

Moore, P. (2015, June 4). *Inbox Zero: Most hate having unread e-mails, a few let them pile up | YouGov*. https://today.yougov.com/topics/lifestyle/articles-reports/2015/06/04/inbox-zero

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2008). Normative Social Influence is Underdetected. *Personality and Social Psychology Bulletin*, *34*(7), 913–923. https://doi.org/10.1177/0146167208316691

Ramli, U. (2021). Social Norms Based Eco-Feedback for Household Water Consumption. *Sustainability*, *13*(5), 2796. https://doi.org/10.3390/su13052796

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, *95*(3), 355–386. https://doi.org/10.1037/0033-2909.95.3.355

Sanguinetti, A., Dombrovski, K., & Sikand, S. (2018). Information, timing, and display: A design-behavior framework for improving the effectiveness of eco-feedback. *Energy Research and Social Science*, *39*, 55–68. https://doi.org/10.1016/j.erss.2017.10.001

Schultz, P. W., Estrada, M., Schmitt, J., Sokoloski, R., & Silva-Send, N. (2015). Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms. *Energy*, *90*, 351–358. https://doi.org/10.1016/j.energy.2015.06.130

Schultz, P. W., Messina, A., Tronu, G., Limas, E. F., Gupta, R., & Estrada, M. (2016). Personalized Normative Feedback and the Moderating Role of Personal Norms: A Field Experiment to Reduce Residential Water Consumption. *Environment and Behavior*, *48*(5), 686–710. https://doi.org/10.1177/0013916514553835

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science*, *18*(5), 429. https://doi.org/10.1111/j.1467-9280.2007.01917.x

Sønderlund, A. L., Smith, J. R., Hutton, C., & Kapelan, Z. (2014). Using Smart Meters for Household Water Consumption Feedback: Knowns and Unknowns. *Procedia Engineering*, *89*, 990–997. https://doi.org/10.1016/j.proeng.2014.11.216

Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E., & Staake, T. (2019). Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nature Energy*, *4*(1), 35–41. https://doi.org/10.1038/s41560-018-0282-1

Webb, T. L., Chang, B. P. I., & Benn, Y. (2013). 'The Ostrich Problem': Motivated avoidance or rejection of information about goal progress. *Social and Personality Psychology Compass*, *7*(11), 794–807. https://doi.org/10.1111/spc3.12071

Appendix

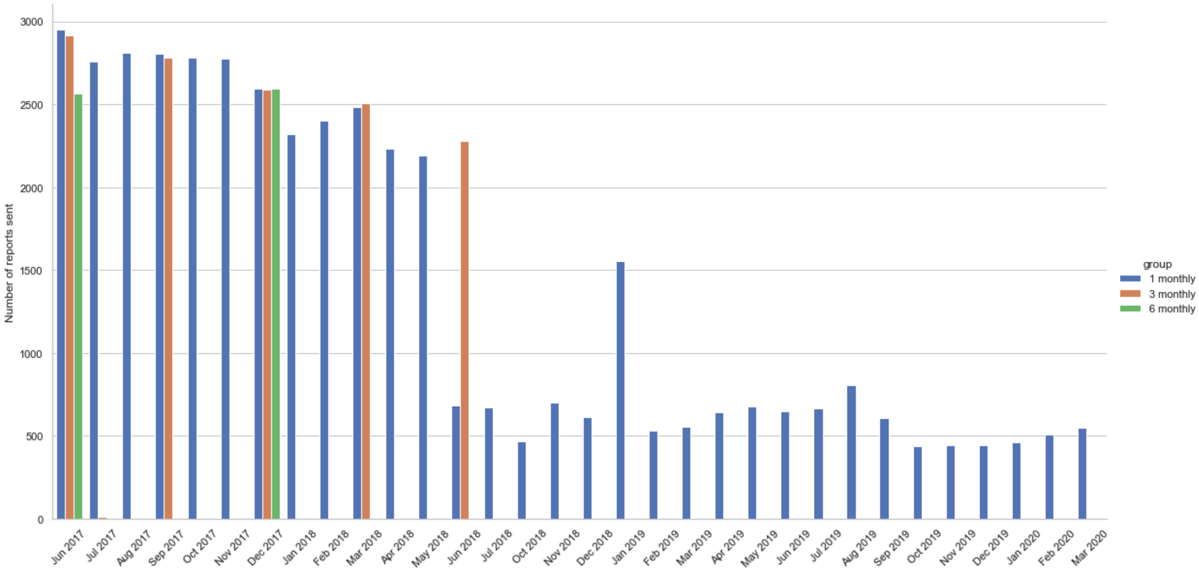

Figure A 2 The time and number of reports sent each month for the different treatment groups for the Frequency experiment.

$$Q_{lpd|Ti}(\tau) = \beta_0 + \ \beta_{Treatment}T_i \ (6)$$

For quantiles 0.5 to 0.95 in increments of 0.1, the following model is fit to each quantile separately, where τ is a representative quantile, and the quantile is an inverse cumulative distribution function. And T_i denotes treatment group assignment.

$$LPD_{it} = \ T_i\pi_m + LPDpre_{it} + \epsilon (7)$$

Conditional Average treatment effect. LPD_{it} as water consumption in litres per day per household. T_i indicator of treatment or control. π_m indicator of month by year. $LPDpre_{it}$ as water consumption in the baseline period.

Experiment/Group	Baseline consumption	Attrition: # treatment months	Attrition: binary of complete treatment months
Frequency (year one)			
1 monthly	-6.305	0.1673	-0.0002
	(4.563)	(0.104)	(0.001)
3 monthly	-5.3613	0.0805	0.0005
	(4.617)	(0.104)	(0.001)
6 monthly	-2.330	0.1774	0.0009
	(4.400)	(0.105)	(0.001)
3 monthly = 1 monthly	-0.944	0.0870	0.001
	(4.682)	(0.104)	(0.001)
6 monthly = 1 monthly	-3.975	0.345***	0.001
	(4.468)	(0.106)	(0.001)
6 monthly = 3 monthly	-3.032	0.258	0.000
	(4.522)	(0.105)	(0.001)
Medium			
Email	1.834	0.071	-0.004
	(2.012)	(0.059)	(0.006)
Paper	-1.213	0.089	-0.006
	(2.021)	(0.058)	(0.006)
Email = Paper	-3.047	0.018	-0.001
	(2.322)	(0.067)	(0.007)

Table A 3 Balance checks for attrition and baseline consumption. The attrition by number of months are t-tests using number of treatment period months available as the dependent variable. The binary attrition is created by categorising households based on whether

6. Paper 3: Double trouble: Concurrently targeting water and electricity consumption using normative messages in the Middle East

Abstract: Personalised normative messages have been shown to be effective at encouraging both electricity and separately water savings. As use of this approach to promote resource savings becomes increasingly widespread, an important question is whether providing such feedback on consumption of the two resources together can yield reductions in both areas. In a field experiment with over 200,000 households in the Middle East, we send households personalised normative messages regarding both their water and electricity consumption on a monthly basis. This intervention saw a statistically significant reduction of around 1.2% for electricity but not for water consumption. Furthermore, we test different ways of concurrently presenting normative messages of both water and energy, including presenting it as a combined eco score. Local treatment effects of these were around 1.2% reduction. Our findings contribute towards nexus thinking around how (not) to concurrently achieve energy and water savings using normative feedback.

6.1. Introduction

The past decade has seen the widespread diffusion of technologies that collect fine-grained, and in some cases real-time, data on consumption of two critical resources: energy and water (Avancini et al., 2019; Mudumbe et al., 2015). Given the environmental significance of these resources, the question of how best to communicate this information to end-users, including households and businesses, in order to encourage them to reduce their consumption is important. A wealth of existing research has examined the effectiveness of feedback. Overall, this literature suggests that feedback can bring about energy and water savings but that its impact varies according to features of the feedback, such as comparison type, frequency and delivery mode (Karlin, Zinger & Ford, 2015; Liu & Mukheibir, 2018).

Among the most well-evidenced forms of feedback in the environmental domain is personalized normative feedback, including both descriptive and injunctive social norms (Schultz et al., 2007; Allcott, 2011; Ayres et al., 2013; Allcott & Rogers, 2014; Ferraro & Price, 2013; Schultz et al., 2019; Wang et al. 2021). Descriptive norms provide target populations with information about their consumption relative to relevant others, like neighbours, with the aim of encouraging them to conserve. Injunctive social norms provide messages that communicate the perceived levels of approval or disapproval of relevant others.

A seminal paper combines these two forms of messages to target reductions in energy behaviours (Schultz et al., 2007). The results indicate that while descriptive norms messages are effective at reducing consumption of energy among high consumers, they give rise to a boomerang effect among people with low energy consumption at baseline. The paper also demonstrates that the boomerang effect can be undone by adding an injunctive message which signals approval of the performance of those low energy using households. This work serves to highlight the potential of normative messages to promote conservation behaviours, as well as their potentially heterogeneous effects across populations.

Off the back of the results of initial studies in this area, there has been a proliferation of utility companies and other organisations (e.g., OPower in the US) targeting energy or water consumption using interventions involving personalized normative feedback (Schultz et al., 2018; Wolske et al., 2020). The companies typically send home resource reports which include both normative feedback as well as conservation tips and other information about energy or water use. Where these efforts have been robustly evaluated, they have tended to provide further evidence of the effectiveness of such reports at encouraging resource conservation (Allcott, 2011; Allcott & Rogers, 2014; Schultz, Javey & Sorokina, 2019). Other work has explored the effectiveness of personalised normative messages in other environmentally significant domains including recycling (Schultz, 1999), the uptake of green technologies (Haffner et al., 2019; DellaValle & Zubareyva, 2019) and support for carbon capture technology (Wang et al. 2021).

Existing research indicates personalised normative interventions targeting resources in one area of consumption can have knock on effects in other environmentally significant consumption areas (Jessoe et al., 2017; Carlsson et al., 2020). For example, Jessoe and colleagues (2021) find that sending households home water reports including normative messages induces a 1.3 to 2.2% reduction in electricity use in summer months. Further analysis indicates that just over a quarter of these reductions were attributable to the indirect reductions in electricity consumption via water savings, suggesting that the messages encouraged electricity saving behaviour as well. Even larger positive spillovers from norms based messages targeting water into energy consumption were documented by Carlsson and colleagues (2020). In contrast to this, Tiefenbeck and colleagues (2013) found that households that received weekly feedback on their water consumption, reduced their water consumption but at the same time increased their electricity consumption. The researchers attributed this finding to moral licensing whereby the moral credit a household experiences when reducing their water consumption licensed them to increase their electricity consumption. Such spillovers have important implications for the cost-effectiveness and attractiveness of home resource report based interventions (Sanguinetti, et al. 2020).

As the literature on this topic has developed, researchers have begun to ask questions about how to optimise the delivery of personalised normative feedback. For example, examining whether coupling it with other interventions like commitments (Jaeger & Schultz, 2017) or incentives (Pellerano et al., 2016; Dolan & Metcalfe 2015) makes them more effective, as well as examining the role of the delivery mode (Dolan & Metcalfe 2015; Schultz et al., 2014). The research has also provided further insights into the differential effects of personalised normative messages across different groups and contexts. For example, descriptive norms based messaging is found to be far less impactful on the consumption of political conservatives compared to liberals (Costa & Kahn, 2013) and on residential energy consumption in Germany compared to the US (Andor et al. 2020).

Both academic and policy discourse surrounding household resource consumption is placing increasing emphasis on 'nexus thinking' (Foden et al., 2019), i.e., accounting for the linkages between environmentally significant consumption across multiple domains including water, energy and food. Given the now widespread prevalence of personalised norm interventions in the environmental space and the evidence of spillovers between resource areas, an important question is how to combine normative messages relating to the consumption of different resources in order to yield the greatest levels of conservation across domains and maximise the messages' environmental benefits. This question is of relevance both in cases where utility companies have the opportunity to collaborate and align their normative based feedback and in situations where, as is commonly the case in the Middle East, utility companies provide services relating to more than one resources, e.g. water and energy or energy and waste.

Against this backdrop, in the current study we carry out a large-scale field experiment with over 200,000 households to examine whether providing both energy and water based personalised normative feedback via email can achieve savings in both domains. The study was carried out in a Middle Eastern metropolis in conjunction with a state-run utility company who supplies both electricity and water to its customers. The metropolis is characterised by high levels of both electricity and water consumption making the site an interesting test case for potentially achieving substantial environmental benefits through personalised norm interventions. Looking at the effectiveness of combining personalised messages on electricity and water provides insights into the feasibility of concurrently encouraging pro-environmental behaviours using this intervention strategy.

In addition to investigating the overall effectiveness of presenting personalised norms based feedback on consumption of both electricity and water, we also examine three different ways of presenting the information: two frames which present the norms based message for each area of consumption separately either sequentially side-by side or with a shared x-axis forming a wing style format, or in a combined eco-score which is computed based on households' relative performance in both areas of consumption (see Figures 13-15). Examining the different presentations of the personalised norms based information speaks directly to the literature around how to optimise the delivery of normative messages. Existing research into the relative effectiveness of different level goals indicates that high-level goals are less effective than low-level goals in promoting energy-saving and other pro-environmental behaviors (Moussaoui & Desrichard, 2016). At the same time, results in the spillovers literature suggest that thinking about pro-environmental goals in concrete rather than abstract terms results in less behavioral consistency (Truelove et al., 2016). Furthermore, Conway and Peetz (2012) found that, in line with construal level theory, temporally close behaviours are perceived more concretely as sub-goals and are therefore more prone to negative spillovers. We therefore propose three main hypotheses. First, social norms based feedback will reduce consumption when compared to the control. Second, the combined eco-score, which can be perceived as a more abstract goal, may be less effective at reducing both electricity and water consumption. Third, that the combined eco-score, as an abstract goal, would be more effective at reducing negative spillovers, compared to delivering feedback to both electricity and water separately. However, due to limitations in the experimental design, it is impossible to measure spillovers specifically by only delivering feedback on one resource and not the other, and so our hypothesis is limited to seeing whether the combined eco-score is able to reduce both water and electricity consumption when compared to delivering feedback of both resources separately.

Overall, the results indicate that personalised normative feedback does yield electricity, but not water, savings when compared to a control group that did not receive this information. When we analyse the electricity consumption of those people who opened the email with the personalised normative messages, we see an annual overall reduction of 1.21%[1] over the course of a 12-month period. This is lower than reductions documented in other high consumption contexts such as the US (Alcott, 2011), which may be attributable to context specific features or to features of the treatment frames. That there is no significant impact on water indicates that, at least in the current study context, delivering normative messages on both water and electricity does not deliver reductions in both areas of consumption. Among the potential explanations for these findings include that the combined message focuses attention on electricity at the expense of water, that the complexity of the message may undermine its effectiveness, or other contextual factors relating to water make reductions harder to achieve in this domain. Further work which compares the relative impact on normative feedback on each domain in isolation to that of the combined feedback is required to explore these potential reasons.

Disaggregating the impact across the different information frames, we find that all three frames, side-by-side, wings and the combined eco-score, yielded significant impacts on electricity but not water. Although the greatest reductions were found in the case of wings, there were no significant differences between the treatment groups. In other words, the framing of the information, at least in the ways explored here, did not influence its impacts. Overall, the results present cautionary evidence against concurrently sending personalised normative messages relating to multiple target behaviours.

In what follows, Section 2 will present the study context and data, Section 3 will present the estimation and results, and Section 4 will discuss the findings and avenues for future research.

6.2. Study context and data

The study was carried out in a Middle Eastern metropolis in collaboration with its nationally owned utility company that provides both energy and water to the population. To be eligible for the study, households had to meet the following criteria, 1) the customer account had to have an active account, 2) the household needed to have at least one month's worth of consumption

[1] This is based on taking the local average treatment effect of all three treatments combined (Table 6) and dividing it by the electricity consumption of the control group at baseline (Table A1).

data, 3) the household had their cooling provided by the same the utility company, 4) the customer did not have multiple accounts, 5) the household did not consume more than 40,000 kilowatt hours per day (kWh/day) of electricity and/or 200,000 imperial gallons of water. Given that an average household in this region consumes 220-360 kWh/year, it is fair to assume that 40,000 kWh/day are either erroneous meter reads or commercial properties. A sample of 218,737 households that met the criteria were selected for the study. The selected households were then randomised into one out of the four experimental groups, three treatment groups and one control. The metropolis has a population of around 3 million, with an average household occupancy of 4.2, which means our sample represents around 30% of the population. As with many countries in the region, both electricity and water are heavily subsidised, which is part of the motivation of the local government owned utility to promote more efficient consumption of water and electricity, as they believe that the subsidising utility bills have caused wasteful consumption amongst the population. The average bill amount based on the average consumption of the sample here is around USD$217.27, with electricity costing USD$136.84 and water costing USD$ 80.43.

The experimental design closely follows the same procedure as previous similar experiments (e.g., Allcott, 2011). Each household in the treatment group was sent an email on a monthly basis that included a link to a portal that contained a consumption horizontal bar graph that compared each household's water and energy consumption against the consumption of their neighbours. Neighbours are defined as households within a geographical area with similar occupancy and house type. Email was chosen as the medium for delivering the treatment by the utility company to avoid the high cost of paper, printing and postage, as well as to reduce the environmental impact of the programme. Email has been found to produce lower treatment effects compared to paper (Dolan & Metcalfe, 2015; Allcott, 2011), but by tracking whether an email was opened, we are able to identify the local average treatment effect. In each treatment group, households saw a different design of the neighbour comparison bar graph, as shown in Figures 13-15. The first treatment group saw two sets of bar graphs side-by-side with electricity on the left and water on the right. The second treatment group - wings - was a similar design but with the graphs sharing an x-axis that is adjacent to each other. The third treatment group was a single bar graph that displayed a consumption score that combined both electricity and water consumption. The score is a standardisation of the two metrics combined. Aside from the neighbour comparison bar graph to deliver the descriptive social norm, a set of smiley faces were also included to highlight an injunctive norm. This serves the purpose of preventing a 'boomerang' effect where high performing households end up licensing themselves to increase their consumption (Schultz et al., 2007). Finally, a set of water and electricity conservation tips were also included in the reports. These tips were refreshed every month with new content.

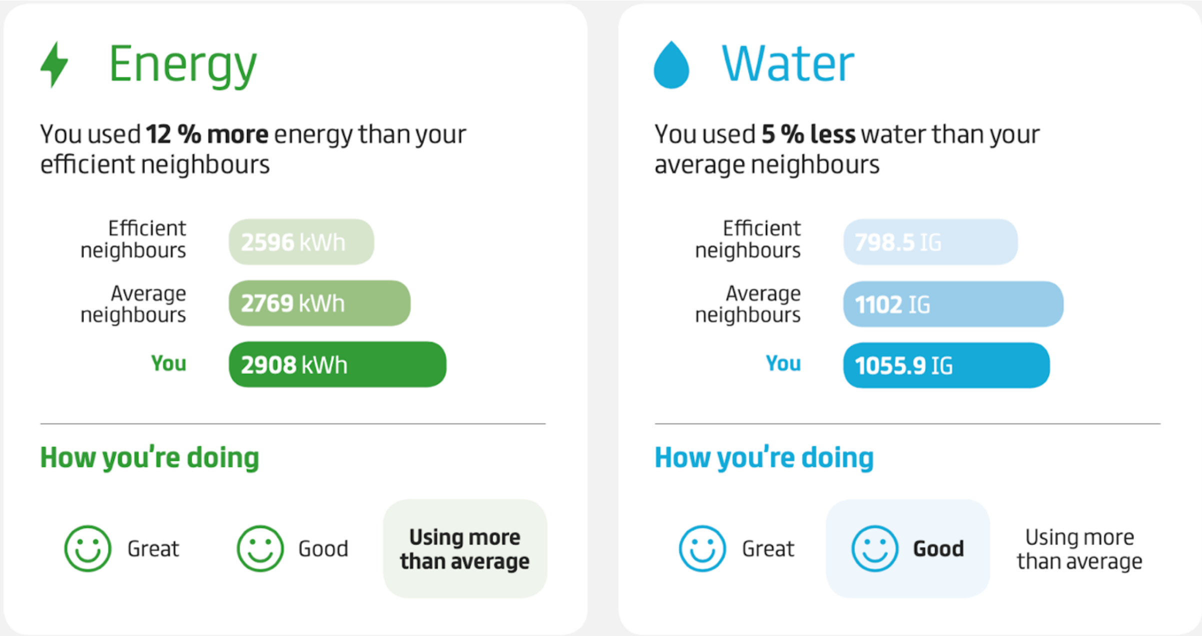

Figure 13 Example of side-by-side treatment group message

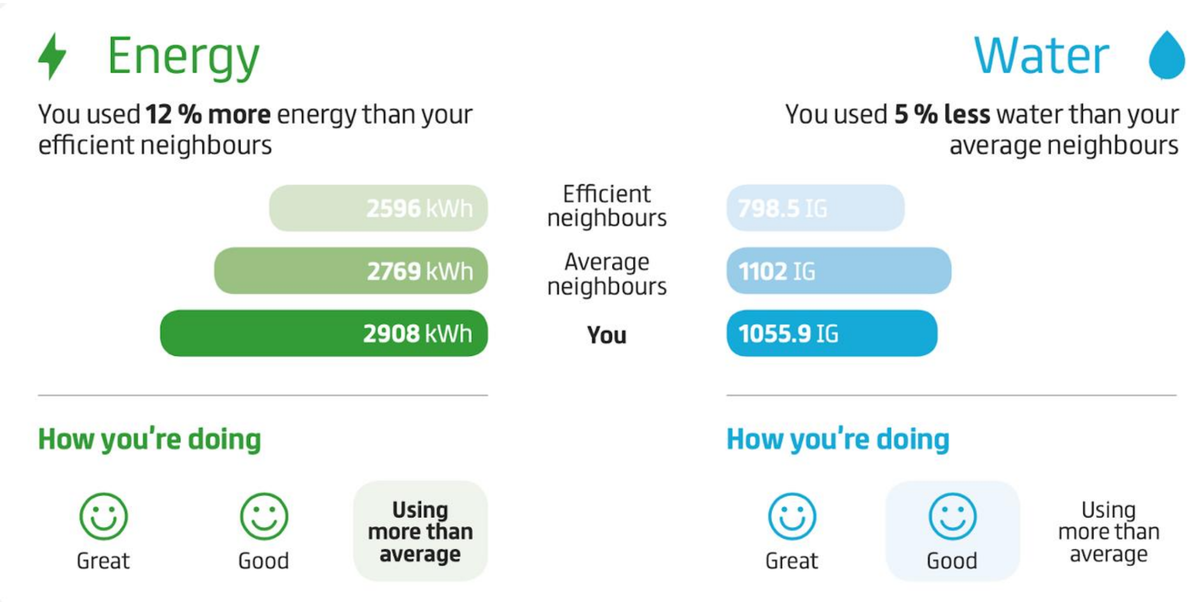

Figure 14 Example of wings treatment group message

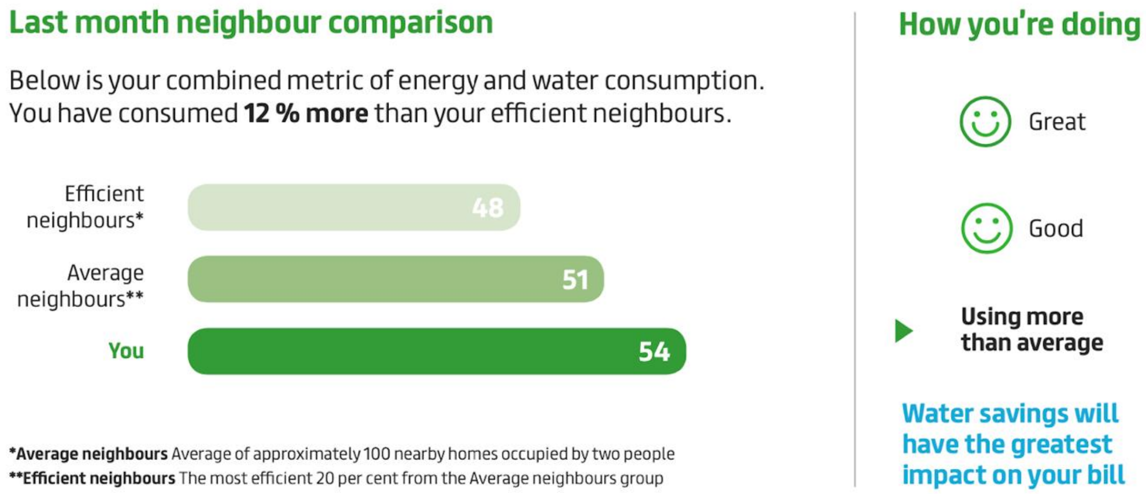

Figure 15 Example of eco-score treatment group message

In order to assess the impact of the treatments, consumption data for both water and electricity was recorded on a monthly basis over a 12 month baseline period (March 2017 to March 2018), and for a subsequent 12 months (March 2018 to March 2019) when households in the treatment group received monthly emails of the treatment. The dependent variables, water and electricity consumption are analysed as imperial gallons per day and kilowatt/hour per day respectively. 3,279 of those meter readings had negative values, suggesting an error in the read, and were therefore dropped from the dataset. Out of the initial sample of 218,737 households, 50,738 households did not have the full 12 months of baseline data available. Missing baseline data was imputed by taking the closest adjacent month's consumption data. For example, if a household had data for January but not for February, then February's baseline was imputed using January's data. This method was preferred over the use of the mean, median or multiple imputation because it better preserved the seasonal differences in consumption. Missing baseline data was balanced between all groups. Furthermore, 9,928 households were missing endline data and were dropped from the dataset. The final dataset consisted of 208,809 households all with positive reads and 12 months' worth of baseline data. Both electricity and water consumption at baseline were balanced across all groups, and can be observed in Figure 16 for electricity and Appendix Figure 3 for water. Descriptive statistics can be seen in Appendix Table A4.

6.3. Estimation & Results

6.3.1. Average treatment effect

We begin by examining the impact of being assigned to any treatment versus being in the control group – the average treatment effect (ATE). We model electricity, and separately water consumption Y_i, conditional on being treated T_i. In Model I, for both electricity and water, we estimate the simple relationship between treatment and consumption (Equation 8). In Model II, we include a vector of controls including house type and whether the household occupants are local or foreign C_i, as well as month fixed effects M_i and baseline consumption B_i (Equation 9).

$$Y_i = \beta_1 T_i + \varepsilon_i \qquad (8)$$

$$Y_i = \beta_1 T_i + C_i + M_i + \beta_2 B_i + \varepsilon_i \qquad (9)$$

We estimate these models using Ordinary Least Squares Regression (OLS) with standard errors clustered at the household level. The results indicate that assignment to treatment does not significantly impact water consumption but leads to reductions in people's electricity consumption by 0.33%[2] over the 12 month period.

We then go on to model the impact of the three different treatment groups. The model is specified as above except that T_i now indicates which of the three treatment groups the participants were assigned to: side-by-side, wings or eco-score. Here we see that of the three treatments, only the wings treatment significantly reduced electricity consumption on average over the 12 month period. None of the treatments impact water consumption.

$$Y_i = \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + C_i + M_i + \beta_4 B_i + \varepsilon_i \qquad (10)$$

Looking longitudinally (Figure 16), we see that the impact of the wings treatment was greatest in the initial period which coincided with the Spring/Summer months, with the treatment remaining significant but decreasing later in the year. For a graph showing the longitudinal estimates for water please see Appendix Figure A3.

[2] This was calculated by taking the -0.191 kWh ATE from the OLS of all three treatments combined (Table A5) and dividing it by 57.30 kWh, the average daily electricity consumption of the control group at baseline (Table A4).

	Model I	Model II	Model III
	ATE Electricity	ATE Electricity	2SLS Electricity
Side-by-side	-0.261 (0.450)	-0.129 (0.104)	-0.467*** (0.194)
Wings	0.010 (0.456)	-0.302*** (0.106)	-1.099*** (0.199)
Combined Score	0.148 (0.458)	-0.146 (0.105)	-0.530*** (0.197)
Side-by-side = Combined	-0.409 (0.458)	0.011 (0.105)	0.103
Wings = Combined	-0.138 (0.464)	-0.158 (0.107)	7.960***
Wings = Side-by-side	0.271 (0.456)	-0.169 (0.106)	10.031***
Baseline	No	Yes	Yes
Time FE	No	Yes	Yes
House Type	No	Yes	Yes
Foreign vs local	No	Yes	Yes
Number of observations	2,163,393	2,163,393	2,163,393
R2	0.000	0.904	0.904

Table 6 OLS and 2SLS analysis of electricity consumption. For Model specification I and II, coefficients from pairwise t-test to test for equality of coefficients were included to show whether the coefficients of the treatment groups were statistically significantly different from each other. For Model specification III, coefficients from Wald test for equality of coefficients were also included. *p<0.1, **p<0.05, ***p<0.001

	Model I	Model II	Model III
	ATE Water	ATE Water	2SLS Water
Side-by-side	1.063 (2.088)	0.641 (0.931)	2.3232 (1.602)
Wings	0.0603 (2.076)	-0.577 (0.909)	-2.102 (1.591)
Combined Score	0.3528 (2.090)	0.069 (0.912)	-0.252 (1.599)
Side-by-side = Combined	0.710 (2.126)	0.710 (0.943)	2.503
Wings = Combined	-0.413 (2.114)	-0.508 (0.920)	1.315
Wings = Side-by-side	-1.123 (2.11)	-1.218 (0.939)	7.463***
Baseline	No	Yes	Yes
Month FE	No	Yes	Yes
House Type	No	Yes	Yes
Foreign vs local	No	Yes	Yes
Number of observations	2,163,393	2,163,393	2,163,393
R2	0.000	0.680	0.680

Table 7 OLS and 2SLS analysis of water consumption. For Model I and II, coefficients from pairwise t-test to test for equality of coefficients were included to show whether the coefficients of the treatment groups were statistically significantly different from each other. For Model specification III, coefficients from Wald test for equality of coefficients were also included. *p<0.1, **p<0.05, ***p<0.001

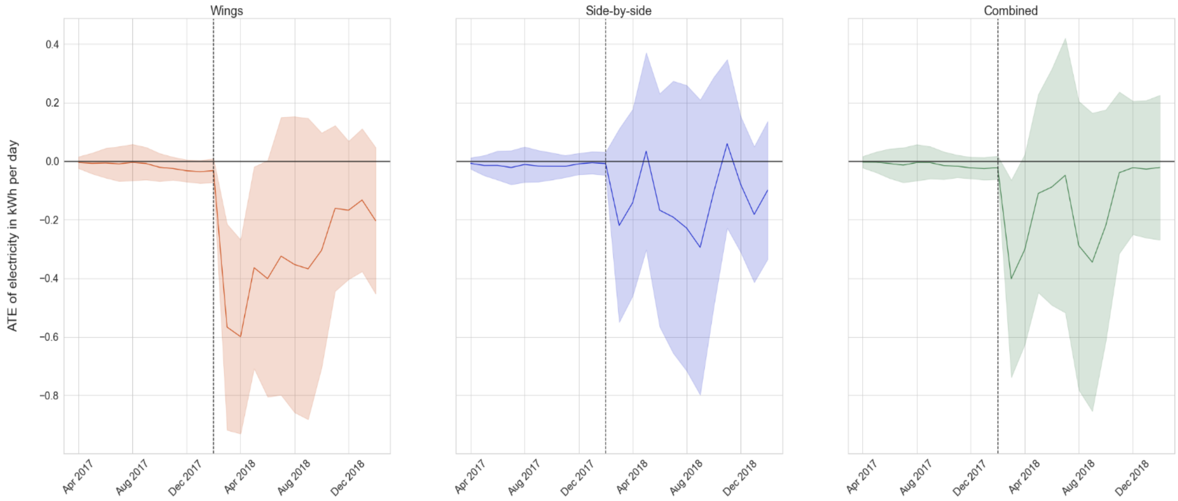

Figure 16 Average treatment effect on electricity over time. The vertical dash line represents the start of the treatment period. The y-axis represents the coefficients are generated from an interaction between group assignment and every month in the pre and post intervention period on electricity consumption. Error bars are 95% confidence intervals.

6.3.2. Quantile regression analysis

To understand the heterogeneity of the treatment effect, we run unconditional quantile regressions for each treatment group separately (Appendix Equation 10). As can be observed in Figure 17, the treatment effect on electricity consumption is heterogeneous for all three treatments, in a largely similar way. Those below the 50th percentile are not moving at all and is a precise zero effect. Between the 50th and 90th percentiles, there is a steady increase in electricity consumption. The average treatment effect seems to be carried by those above the 90th percentile, although this is difficult to discern due to the noise. Overall, these quantile distributions indicate that all three treatments have very similar impacts. This suggests that the significant effect found in wings does not necessarily reflect systematic differences between the different treatments, but could simply be a statistical artefact of the sample. This explanation is further supported by pairwise t-tests of equality of coefficients presented in Table A 3. This is an example of quantile regressions being able to better interpret ATEs as the distributional impact of the wings treatment group looks almost identical to the others. This uniformity can also be observed for water consumption, although with mostly a precise zero effect, in Appendix Figure A4.

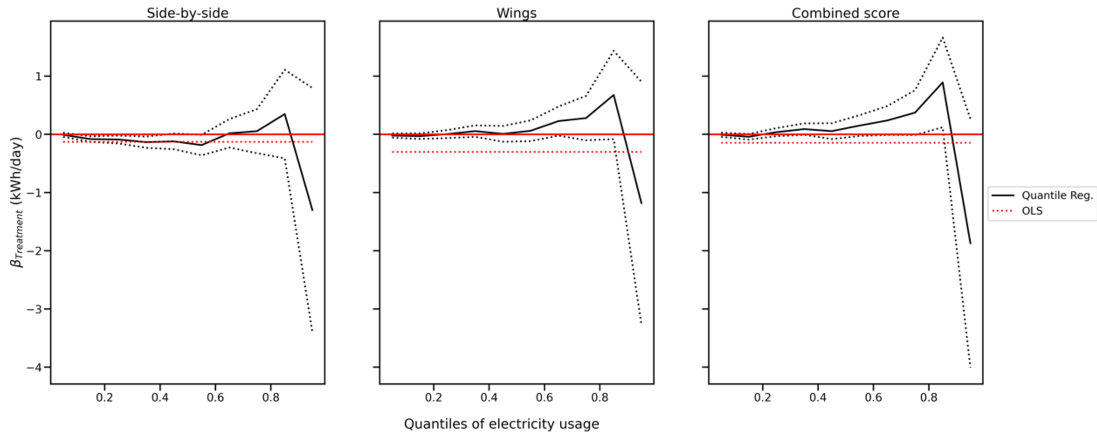

Figure 17 Unconditional quantile regression of treatment effects on electricity. The red dashed lines represent the ATE based on Model I. The solid black line represents the treatment effect of each quantile. Therefore, no variables were included as control here. Confidence intervals are represented as black dashed lines. Higher deciles reflect higher baseline consumption

6.3.3. Conditional average treatment effect

The treatment effects of these feedback interventions are rarely homogenous, and one of the main sources of heterogeneity is based on a household's baseline consumption (Brent et al. 2020; Ferraro & Miranda, 2013). We therefore look at the conditional average treatment effect (CATE), conditional on deciles of baseline consumption, to determine whether the treatment effect is greater for households with higher baseline consumption (Appendix Equation 11). We disaggregate average treatment effects by deciles based on a given household's mean consumption over the baseline period, and interact this with treatment group assignment within the same model specification used for determining the ATE. See Figure 18.

Across all three CATE estimates the majority of the treatment effect occurs in the top decile with all other deciles having either very small or no effect at all. Households in the 8th and 9th deciles also seem to increase consumption, which may explain the overall small ATE for wings and the null effects for the other treatment groups. As a point of reference, the average daily consumption of households in the top decile is 251.9 kWh. The similarity in the pattern of distribution between these CATE figures and the quantile distribution is suggestive of rank invariance, which is to say that households are not swapping ranks over the distribution over time nor as a result of treatment. CATE of water consumption is available in Figure A5.

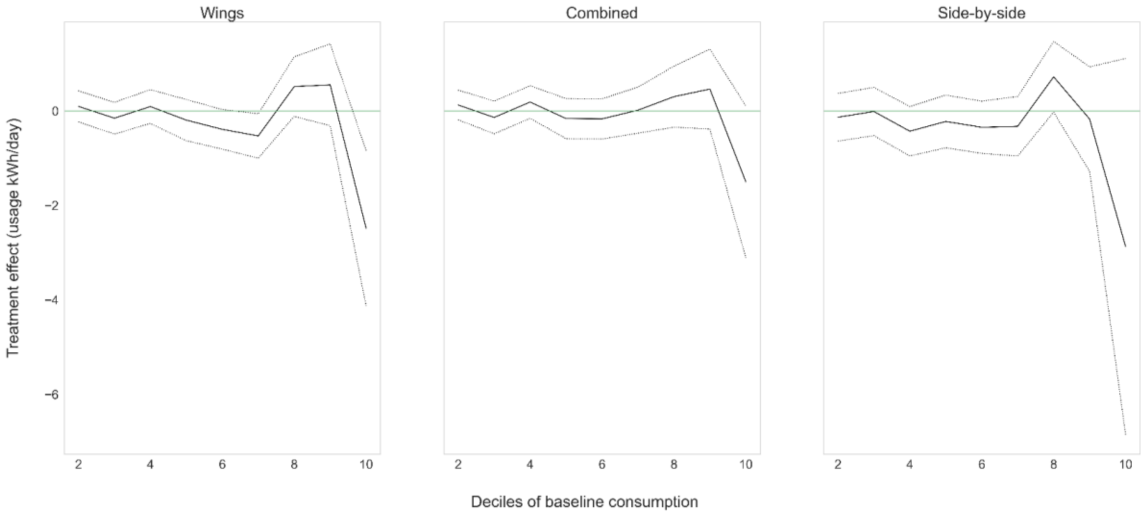

Figure 18 Conditional average treatment effect (solid lines) based on an interaction between deciles of baseline electricity consumption and treatment group assignment. Decile 10 indicates households with the largest amount of electricity consumption during the baseline period. Decile 1 is omitted as the comparison decile. Error bars (dashed lines) are 95% confidence intervals.

6.3.4. Local average treatment effect

Importantly, being randomly allocated to be in a treatment group does not automatically ensure that individuals were exposed to the feedback on their consumption in a given month as they may not have opened the email. In fact, of the total emails sent, on average 27.51% of treatment emails were opened per month over the course of the one year study[3]. We therefore follow up our ATE analysis by adopting a two stage least squares approach (2SLS), using opening the email of the treatment in each month of the experimental period as the instrument to estimate the local average treatment effect (LATE) (Imbens and Angrist, 1996). These models take the following form:

$$First\ Stage\colon T_i = \rho_1 Z_i + \rho_2 C_i + \rho_3 M_i + \rho_3 M_i + \rho_4 B_i + \varepsilon_i \qquad (11)$$

$$Second\ stage\colon Y_i = \Gamma_1 T_i + \Gamma_2 C_i + \Gamma_3 M_i + \Gamma_4 B_i + \varepsilon_i \qquad (12)$$

T_i is the endogenous variable, whether person i opened the email, and $\rho_1 Z_i$ is the instrumental variable of treatment group assignment. Again, we included a vector of controls including house

[3]Due to technical errors in the data storing of the email read receipts, 25% of the data of those read receipts were missing. To be conservative, we converted those missing values as ‘not open email’.

type and whether the household is local or foreign C_i as well as monthly fixed effects M_i, and baseline consumption B_i See Model III in Table 6 and 7 for results relating to electricity and water consumption respectively. Examining the impact of the feedback on consumption for those who opened the email across the two resources, here again we see that although the treatments significantly impact electricity consumption, there is still no effect on water consumption.

Moving on to the disaggregated results, we see that while all three feedback frames reduce electricity, none have a significant impact on water (including when adjusting for testing multiple hypotheses using a bonferroni correction). See Figure 19 and Figure A6 of the Appendix for coefficient plots. Of the three treatments the wings design yields the greatest magnitude reductions of -1.096kWh/day (1.9% overall), followed by the combined eco-score with -0.529 kWh/day (0.911%) and side-by-side with -0.465kWh/day (0.797%), with the difference between the impact of the wings treatment and the other two being statistically significant.

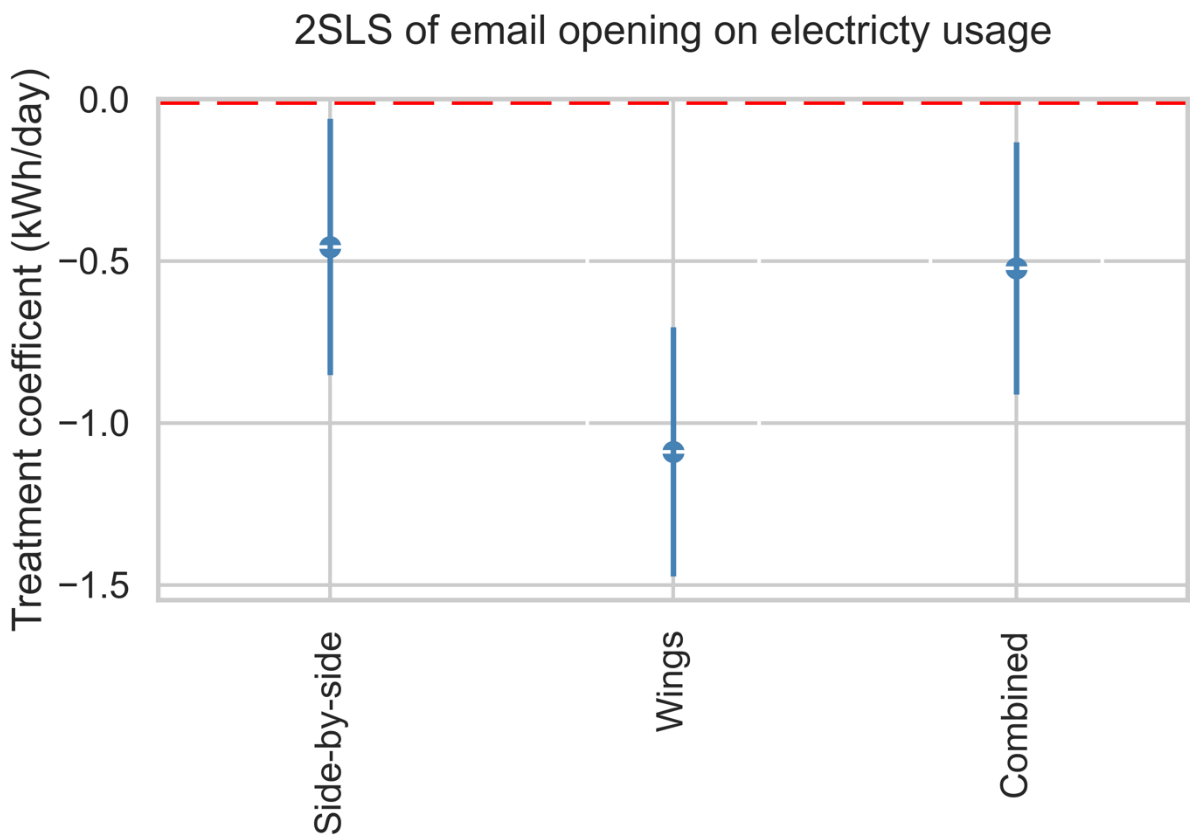

Figure 19 Coefficient plot of electricity usage in kilowatt hours per day from 2SLS for each treatment group, using an interaction between treatment assignment as the IV and monthly email opening as the endogenous variable. Error bars are 95% confidence intervals.

6.4. Discussion

In the current study, we present the first work to explore the effectiveness of concurrently presenting households with personalised normative feedback on two areas of environmentally significant consumption: electricity and water. The intervention delivered a 1.9% reduction in electricity consumption over a 12-month period, when looking at the LATE using email open rate of the wing treatment, or 1.2% when looking at all treatments combined. Importantly, as participants were required to navigate to an online portal from the email to see their feedback and we are not able to monitor the frequency with which people did this, this LATE should be considered lower bounds for the impact of actually seeing the feedback and is likely an underestimate. The percentage change is similar to the treatment effects found in the US where personalised normative feedback delivered as part of paper home energy report letters achieved reductions in electricity consumption of around 2% (Allcott, 2011; Allcott & Rogers, 2014)[4]. The treatment effects found here are also expectedly higher than the 0.7% savings achieved with a similar intervention in a German context (Andor et al., 2020). As suggested by Andor and colleagues (2020) greater savings can be achieved in populations with higher levels of consumption. Although it is interesting to note, in our Middle Eastern context, electricity consumption is three times greater, at 90kWh/day, than that of the US, at around 30kWh/day.

With regards to water, despite equivalent levels of consumption in both the case study metropolis and the US (both of which are high by international standards), water consumption appears to have been unaffected by normative feedback in this context, even when looking at the LATE. Given that the findings of home water reports have been generalisable into other contexts such as the US (Schultz et al., 2018), the UK (Ramli, 2020) and Colombia (Carlsson et al., 2020), it seems that there is something related to the local context that creates friction for households to reduce their water consumption. Research on goal shielding suggests that when individuals have multiple goals, they are prone to concentrate on only one goal. This is understood to be particularly likely to occur when the goals serve the same overarching purpose (Shah, Friedman & Kruglanski, 2002). Even if goal shielding was the reason for the null effect in

[4] Importantly the ~2% documented by Allcott represents the average treatment effect. This is as a result of the study design being unable to identify whether recipients opened the home energy report letter they were mailed or not.

water, it is unclear why conservation behaviour was systematically directed towards electricity rather than water. It may simply be the case that while water consumption in this region is high, households may believe that the entirety of their water consumption is necessary, and that there is no scope for conservation. Alternatively, as water is cheaper than electricity, it may be that household are more financially more motivated to reduce their electricity consumption vs water. These differences in cost are prevalent across most countries however, and so if price was really such a strong motivator, we should be seeing similar null results in studies using social norms on water conservation in other contexts as well. Therefore, there seems to be some effect of presenting both goals together that may be driving the null result for water, or some unique cultural factor interacting with the effect. Future research should examine the differences in providing water feedback alone compared to water feedback paired with electricity to shed further light on this issue, and to explore these interventions in more non-Western contexts.

Other work on the impact of dynamic pricing programs on electricity demand indicates that concurrently offering two forms of dynamic pricing is less effective than only offering one form in isolation, despite the increased incentive involved (Jacobsen & Stewart, 2020). This result highlights that increasing the complexity of strategies aimed at encouraging demand reductions can backfire. Another interpretation of the findings in the current work, therefore, is that the complexity of the normative intervention targeting two resource areas may have diluted its impact. Future research should examine the differences in providing water feedback alone compared to water feedback paired with electricity to shed further light on these issues.

While existing research has documented electricity savings as knock on effects from personalised norm interventions targeting water consumption (Jessoe et al. 2017; Carlsson et al., 2020), as far as we are aware no research to date examined the reverse, i.e., spillovers from interventions targeting electricity consumption on water consumption. While our null findings in relation to water in the current work cannot speak directly to this gap, they do raise the question of whether spillovers into electricity savings from interventions targeting water are more obtainable than those arising in water from interventions targeting electricity. Given the increasing emphasis placed on the interdependencies between consumption of these resources in households – the water-energy-food nexus (Foden et al. 2019) future work should explore the drivers of both direct and indirect reductions from personalised normative messages achieved across all of these domains when one area of consumption is targeted.

Looking at the different forms of combined feedback, we find that according to the LATE for electricity consumption, looking at those that opened the email, the wings treatment has the largest coefficients, and is significantly different from that of the other two treatments. More specifically, the coefficient for the local average treatment effect on the wings treatment is 1.09 kWh/day, while that of the side by side treatment is 0.47kWh/day and that on the combined score is 0.53kWh/day. By way of comparison, the average daily reduction for wings is very close to that of the electricity an iron uses per hour at 1.08 kWh, and for the side by side treatment and the combined score it is closer to half an hour.

Although existing research has found that adding embodied energy feedback to feedback on water consumption yielded significant reductions in water consumption (Jeong et al., 2014), no research to date has examined the impact of presenting water and electricity personalised normative feedback together. Prior to the study, therefore, it was unclear whether combining the information or presenting it separately would best promote overall environmental performance. On the one hand, presenting the information in a combined eco-score makes salient the connection between the two areas of consumption and their relationship to environmental impact. Features which should theoretically limit negative behavioural spillovers and encourage positive ones (Truelove et al., 2014). On the other hand, presenting the feedback separately provides consumers feedback relating to concrete, rather than abstract, goals – something which should promote goal attainment (Moussaoui & Desrichard, 2016). That the wings score achieves larger reductions than the eco score suggests there may be benefits to separating out the feedback. However, as the eco score and the side by side treatments have equivalent impacts, the evidence is rather mixed and perhaps that these relative benefits counteract one another.

In regard to cost effectiveness of the intervention, if the average daily savings based on the ATE of the wings treatment is 0.302 kWh, which is around 9.19 kWh per month, and the cost of electricity in the metropolis is around USD$0.082 per kWh, then the savings per household per month is around USD$0.75. As the treatment itself costs around USD$0.21 per email per month, the intervention appears to be cost effective. Although true cost effectiveness is difficult to determine since households do not pay for these reports, and the benefit to the utility is difficult to calculate as it relates to the operational costs of delivering energy, as well as the subsidies provided to households.

The study design shares limitations with some previous work on normative messages targeting energy and water savings, highlighting directions for future work. First, the study does not examine the underlying behavioural processes causing resource reductions despite their relevance for both theory and practice. Future work could make use of graphical causal models and include surveying household energy- and water-saving measures in order to shed light on the behavioural pathways behind the impact on electricity consumption documented here (Bhushan et al.,2018; Sanguinetti et al., 2020). A second issue is that the study focuses on water and electricity resource consumption as the outcomes of interest without providing insights into the overall consumer welfare effects. In contrast, Allcott and Kessler (2019) elicit consumer willingness to pay for home energy reports in order to explore the welfare impacts on those receiving the intervention. Future work targeting both energy and water could adopt this approach to better understand the value of the intervention from the consumers' perspective.

Furthermore, in order to carry out the current study it was necessary to partner with a utility company. Such partnerships typically require flexibility and mutual benefits (Müftüoglu et al., 2018). Compromises to the research design, in particular in relation to excluding treatment groups in which only single resources were targeted, were necessary in order to carry out the study. Including such groups would have allowed us to causally identify the spillover effect of just targeting one of the resources. This, however, could not be done as it would have meant denying the service of the feedback to a portion of the population. Additionally, despite existing evidence to suggest that paper based messages are more effective than email (Dolan & Metcalfe, 2015; Schultz et al., 2014), it was not possible in the current context to send paper letters and so we proceeded with emails. The environmental and financial cost of sending paper letters to a sample size of this size at that frequency was not seen as sustainable to the utility. Another suboptimal feature of the design was that recipients in the treatment group had to click on a link in the email in order to view the normative messages, creating further barriers to treatment. Despite these practical limitations, the intervention represented a cost-effective solution for the encouragement of electricity reduction, and they have now rolled out the intervention to their entire customer base. They have also subsequently embedded the normative messages into the body of the email that they send. This change will likely enhance the intervention's effectiveness. Future work should examine the relative effectiveness of targeting the two resource areas in isolation compared to together, and further explore the impact of different features of the delivery mode on the impact of normative messages.

Overall, the results of this study indicate that concurrently presenting personalised normative feedback on both electricity and water consumption yields reductions in the former but not the latter. As social norms based messages become more and more widespread there is an onus on utility companies to consider the effectiveness of sending normative messages targeting multiple domains. While existing studies demonstrate that personalised normative messages can be effective at encouraging water and electricity savings when targeted separately, the current results suggests that providing feedback on electricity and water at the same time may lead to a focus on just one sub-goal (electricity) over the other (water).

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics*, *95*(9-10), 1082-1095.

Allcott, H. & Rogers, T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *Am. Econ. Rev.* 104, 3003–3037 (2014).

Andor, M. A., Gerster, A., Peters, J., & Schmidt, C. M. (2020). Social norms and energy conservation beyond the US. *Journal of Environmental Economics and Management*, *103*, 102351.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Avancini, D. B., Rodrigues, J. J., Martins, S. G., Rabêlo, R. A., Al-Muhtadi, J., & Solic, P. (2019). Energy meters evolution in smart grids: A review. *Journal of cleaner production*, *217*, 702-715.

Ayres, Ian, Sophie Raseman, and Alice Shih. "Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage." *The Journal of Law, Economics, and Organization* 29.5 (2013): 992-1022.

Bator, R. J., Phelps, K., Tabanico, J., Schultz, P. W., & Walton, M. L. (2019). When it is not about the money: Social comparison and energy conservation among residents who do not pay for electricity. *Energy Research & Social Science*, *56*, 101198.

Bhushan, N., Steg, L., & Albers, C. (2018). Studying the effects of intervention programmes on household energy saving behaviours using graphical causal models. *Energy research & social science*, *45*, 75-80.

Brent, D.A., Lott, C., Taylor, M., Cook, J., Rollins, K., & Stoddard, S. (2020). What causes heterogenous response to social comparison messages for water conservation? Environmental and Resource Economics, 77, 503-537.

Carlsson, F., Jaime, M., & Villegas, C. (2020). Behavioral spillover effects from a social information campaign. *Journal of Environmental Economics and Management*, 102325.

Costa, D. L., & Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, *11*(3), 680-702.

DellaValle, N., & Zubaryeva, A. (2019). Can we hope for a collective shift in electric vehicle adoption? Testing salience and norm-based interventions in South Tyrol, Italy. Energy Research & Social Science, 55, 46-61.

Dolan, P., & Metcalfe, R. (2015). Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation. *Becker Friedman Institute for Research in Economics Working Paper*, (2589269).

Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, *95*(1), 64-73.

Ferraro, P. J., & Miranda, J.J. (2013). Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment. Resource and Energy Economics. 35. 356-379.

Foden, M., Browne, A. L., Evans, D. M., Sharp, L., & Watson, M. (2019). The water–energy–food nexus at home: New opportunities for policy interventions in household sustainability. *The Geographical Journal*, *185*(4), 406-418.

Hafner, R., Elmes, D., Read, D., & White, M. P. (2019). Exploring the role of normative, financial and environmental information in promoting uptake of energy efficient technologies. *Journal of Environmental Psychology*, *63*, 26-35.

Jacobsen, G., & Stewart, J. I. (2020). How Do Consumers Respond to Price Complexity? Experimental Evidence from the Power Sector. *Experimental Evidence from the Power Sector (December 1, 2020).*

Jaeger, C. M., & Schultz, P. W. (2017). Coupling social norms and commitments: Testing the underdetected nature of social influence. *Journal of Environmental Psychology*, *51*, 199-208.

Jeong, S. H., Gulbinas, R., Jain, R. K., & Taylor, J. E. (2014). The impact of combined water and energy consumption eco-feedback on conservation. *Energy and Buildings*, *80*, 114-119.

Jessoe, Katrina, et al. "Spillovers from behavioral interventions: Experimental evidence from water and energy use." *Journal of the Association of Environmental and Resource Economists* 8.2 (2021): 315-346.

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological bulletin*, *141*(6), 1205

Liu, A., & Mukheibir, P. (2018). Digital metering feedback and changes in water consumption–A review. *Resources, Conservation and Recycling*, *134*, 136-148

Moussaoui, L. S., & Desrichard, O. (2016). Act local but don't think too global: The impact of ecological goal level on behavior. *The Journal of social psychology*, *156*(5), 536-552.

Mudumbe, M. J., & Abu-Mahfouz, A. M. (2015). Smart water meter system for user-centric consumption measurement. In *2015 IEEE 13th international conference on industrial informatics (INDIN)* (pp. 993-998). IEEE.

Müftüoglu, I. B., Knudsen, S., Dale, R. F., Eiken, O., Rajak, D., & Lange, S. (2018). Rethinking access: Key methodological challenges in studying energy companies. *Energy research & social science*, *45*, 250-257.

Pellerano, J. A., Price, M. K., Puller, S. L., & Sánchez, G. E. (2017). Do extrinsic incentives undermine social norms? Evidence from a field experiment in energy conservation. *Environmental and resource Economics*, *67*(3), 413-428.

Sanguinetti, A., McIlvennie, C., Pritoni, M., & Schneider, S. (2020). Beyond Curtailment and Efficiency: Identifying Household Energy and Water Saving Measure Classes.

Schultz, P. W. (1999). Changing behavior with normative feedback interventions: A field experiment on curbside recycling. *Basic and applied social psychology*, *21*(1), 25-36.

Schultz, P. W., Javey, S. & Sorokina, A. Social comparison as a tool to promote residential water conservation. *Front. Water* 1, 1–9 (2019).

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429-434.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2018). The constructive, destructive, and reconstructive power of social norms: Reprise. *Perspectives on psychological science*, *13*(2), 249-254.

Schultz, P. Wesley, et al. Personalized normative feedback and the moderating role of personal norms: A field experiment to reduce residential water consumption." *Environment and Behavior* 48.5 (2014): 686-710.

Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: on the antecedents and consequences of goal shielding. *Journal of personality and social psychology*, *83*(6), 1261.

Tiefenbeck, V., Staake, T., Roth, K., & Sachs., O. (2013). For better or for worse? Empirical evidence of moral licensing un a behavioural energy conservation campaign. Energy Policy 57, 160-171.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127-138.

Wang, T., Shen, B., Springer, C. H., & Hou, J. (2021). What prevents us from taking low-carbon actions? A comprehensive review of influencing factors affecting low-carbon behaviors. Energy Research & Social Science, 71, 101844.

Wolske, K. S., Gillingham, K. T., & Schultz, P. W. (2020). Peer influence on household energy behaviours. *Nature Energy*, *5*(3), 202-212.

Appendix

	Pre	Post	Pre	Post	n	n
	Water (Ig/day)	Water (Ig/day)	Electricity (kWh/day)	Electricity (kWh/day)	Households	Observations
Treatment group						
Control	243.01 (400.19)	241.75 (400.45)	58.09 (91.46)	57.23 (90.48)	52,986	548,174
Side-by-side	243.77 (401.57)	242.81 (414.91)	58.08 (91.17)	56.97 (90.53)	53,010	549,765
Wings	243.54 (401.66)	241.69 (403.36)	58.41 (93.63)	57.24 (91.48)	51,256	531,685
Combined Score	243.30 (403.05)	242.10 (409.84)	58.39 (93.46)	57.38 (92.23)	51,557	533,769
House type						
Apartment	131.63 (121.92)	129.85 (119.26)	29.05 (27.49)	28.37 (26.59)	163,518	1,669,010
Villas	618.33 (682.75)	617.94 (696.08)	156.38 (148.80)	154.22 (146.67)	44,914	490,469
Other	922.86 (996.94)	1005.68 (1100.14)	186.60 (193.91)	195.83 (210.21)	377	3,914
Nationality						

Local	746.72 (800.74)	748.12 (820.32)	171.66 (178.24)	170.58 (176.06)	2,4621	278,554
Foreign	169.02 (217.71)	167.31 (217.78)	41.44 (54.09)	40.45 (52.72)	184,188	1,884,839

Table A 4 Descriptive statistics of each treatment group, subgroup in the pre and post period across water and electricity usage. Balance checks using OLS of baseline consumption and group assignment show no statistically significant differences between the control and treatment groups.

	Model I OLS Electricity (kWh/day)	Model II OLS Electricity (kWh/day)	Model III OLS Water (IG/day)	Model IV OLS Water (IG/day)	Model V 2SLS Electricity (kWh/day)	Model VI 2SLS Water (IG/day)
Effects of all treatments combined	-0.0372 (0.369)	-0.191 (0.085)	0.459 (1.689)	0.005 (0.742)	-0.695 (0.160)	-0.019 (1.423)
Baseline	No	Yes	No	Yes	Yes	Yes
Time FE	No	Yes	No	Yes	Yes	Yes
House type	No	Yes	No	Yes	Yes	Yes
Number of observations	2,163,393	2,163,393	2,163,393	2,163,393	2,163,393	2,163,393
R2	0.000	0.904	0.000	0.680	0.904	0.680

Table A 5 OLS and 2SLS results of all treatment groups combined for water and energy.

Figure A 3 Average treatment effect on water over time. The vertical dash line represents the start of the treatment period. The y-axis represents the coefficients generated from an interaction between group assignment and every month in the pre and post intervention period on electricity consumption. Error bars are 95% confidence intervals.

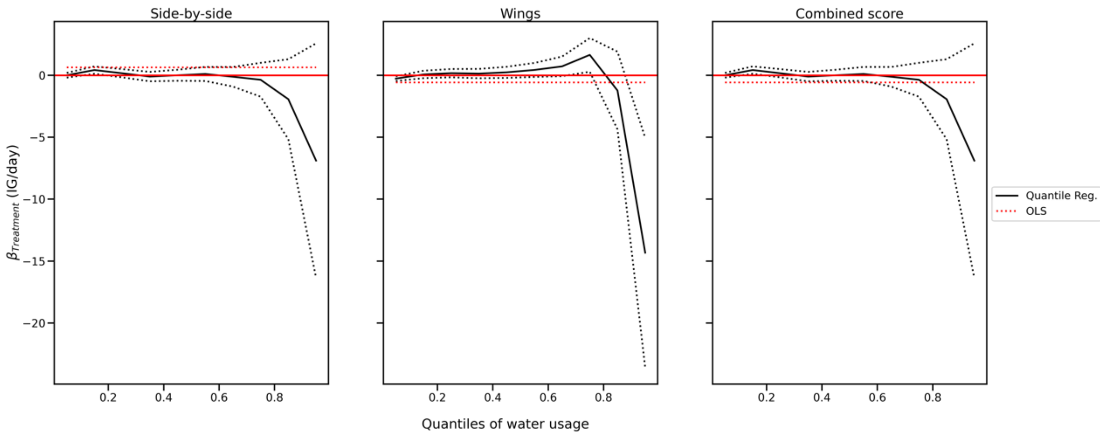

Figure A 4 Quantile treatment effects of water consumption across all three treatments. The red dashed lines represent the ATE based on Model I. The solid black line represents the treatment effect of each quantile. Confidence intervals are represented as black dashed lines. While there may seem like one of the upper quantiles is significant, the size is so small that the effect is likely to disappear once multiple corrections has been implemented.

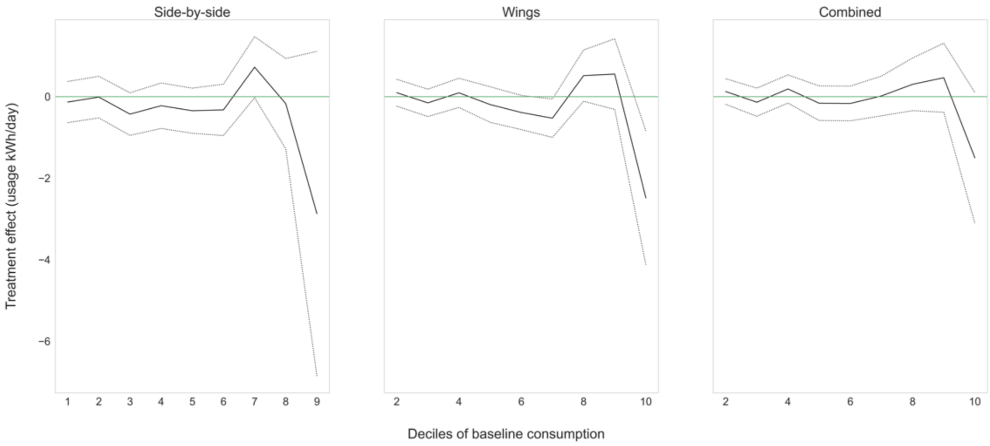

Figure A 5 Conditional average treatment effect (solid lines) based on an interaction between deciles of baseline water consumption and treatment group assignment. Decile 10 indicates households with the largest amount of electricity consumption during the baseline period. Decile 1 is omitted as the comparison decile. Error bars (dashed lines) are 95% confidence intervals.

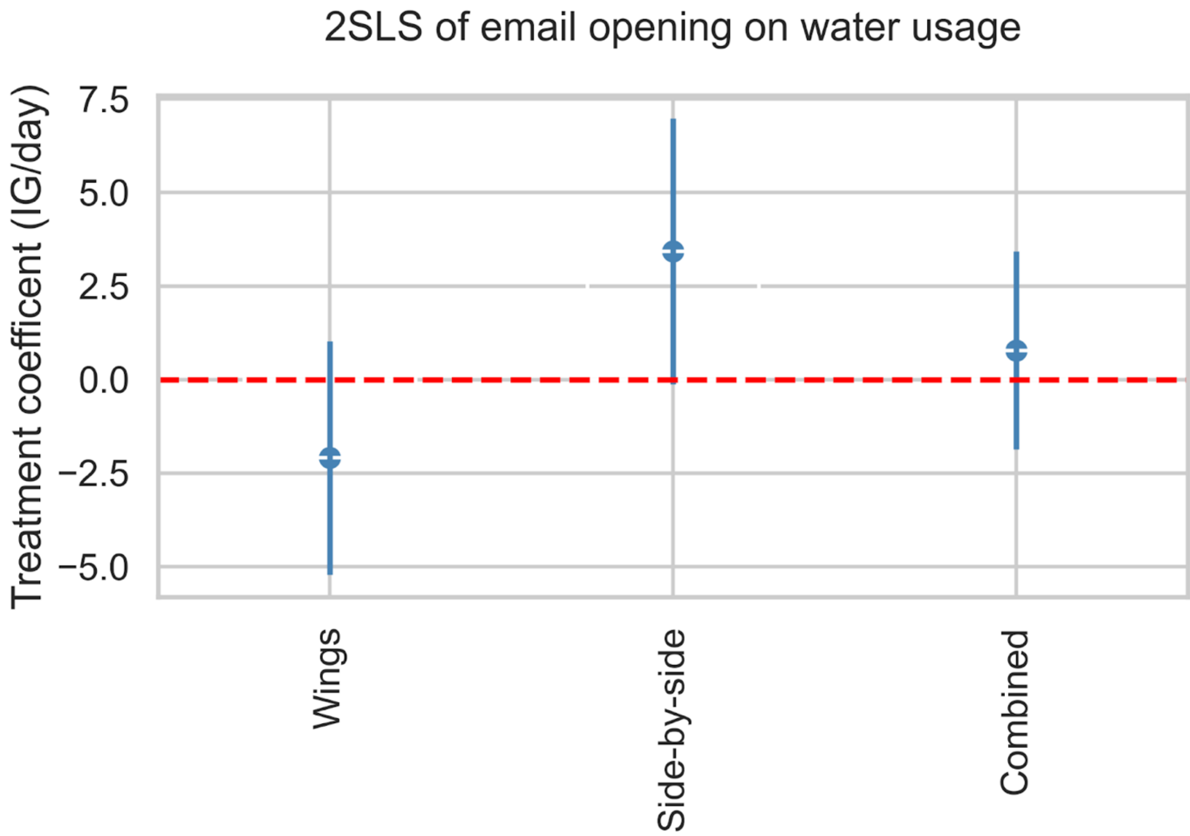

Figure A 6 Coefficient plot of water usage in imperial gallons per day from 2SLS for each treatment group, using an interaction between treatment assignment as the IV and monthly email opening as the endogenous variable. Error bars are 95% confidence intervals.

$$Q_{lpd|Ti}(\tau) = \beta_0 + \beta_{Treatment} T_i \qquad (13)$$

For quantiles 0.5 to 0.95 in increments of 0.1, the following model is fit to each quantile Q separately, where τ is a representative quantile, and the quantile is an inverse cumulative distribution function, and T_i denotes treatment group assignment.

$$Y_{it} = T_i \pi_m + B_{it} + \epsilon \qquad (14)$$

Conditional average treatment effect for electricity. Y_{it} is household consumption in kWh/day or imperial gallons/day. T_i denotes the treatment indicator for household i, while π_m denotes the month, and B_{it} denotes consumption in kWh/day or imperial gallons/day.

7. Conclusion and discussion

Eco-feedback is an important approach to tackling household water and energy consumption, with a view to helping reduce our impact on the environment. As it continues to be widely adopted by governments and companies all around the world, the importance of understanding its design mechanics and how these can affect the overall success of the intervention also grows.

All three papers primarily contributed to the literature of eco-feedback interventions with additional evidence of their effectiveness in reducing household water and energy consumption. Papers 1 and 2 were the first examples of eco-feedback being used to reduce household water consumption in the UK, while paper 3 provided the first example of any eco-feedback being implemented in the Middle East, but also the first example anywhere of eco-feedback being delivered for both water and energy at the same time. The effect sizes found in papers 1 and 2 of around 2% reduction in water consumption were similar to those found in previous studies, while the reduction in electricity consumption of 1.8% in paper 3 was similar to previous studies only when looking at the local average treatment effect.

The effect sizes here seem small in comparison to other behavioural science based interventions, but the small effect sizes found in large scale field experiments such as these are likely to be more reliable than the potentially inflated effect sizes reported in smaller scale lab studies (Mertens, et al. 2022; Szasi et al. 2022; Hummel & Maedche, 2019; Della Vigna & Vignos, 2020). Furthermore, while these effect sizes are small, the intervention benefits from being easy enough to implement to a very large number of households, which allows it to have a cumulative effect. For example, the interventions conducted in papers 1 and 2 achieved water savings of around 42 million litres of water a year[5]. This is comparable to reducing demand for water by an equivalent volume of two reservoirs. This means that water companies may not need to build additional reservoirs to meet growing demand. Reservoirs, which are associated with negative environmental impact. Beyond the cumulative impact of the small effect sizes, eco-feedback could also have larger impact through its ability to make environmental issues more salient. Being aware of pro-environmental behaviour in one area has been shown to increase pro-environmental behaviour in other areas (Evans, et al. 2012). Additionally, being aware of one's own energy consumption is an important factor in promoting pro-environmental choices (Li et al. 2021). Therefore, beyond tackling water or energy consumption alone, eco-feedback could have positive spillover effects into other environmental behaviours. Although the eco-feedback reports may need to be more explicitly framed as pro-environmental to have this effect. Furthermore, in line with theory of cognitive dissonance together with the findings by Fishbach et al. (2006), positive eco-feedback framed as pro-environmental could lead to greater personal commitment to pro-environmental behaviours. While we have seen that eco-feedback in water does lead to positive spillover in energy (Carlsson et al. 2020), future studies could explore the possibility of the positive spillover occurring across other pro-environmental behaviours, especially when the feedback itself is framed as pro-environmental.

Another approach to increasing the overall impact of eco-feedback could also be to combine it with other behavioural interventions. For example, as noted by Nisa et. al (2019), interventions related to choice architecture are more effective than eco-feedback. As discussed in the

[5] The average baseline consumption of the control group across all three experiments was 285 litres/day. A 2% reduction a day is equivalent to 5.7 litres a day, which is 2,080 litres a year. Multiplied across the total sample of all three experiments of around 20,334 households is 42,294,720 litres a year. This is of course a rough estimate as not all treatment groups achieved the same effect size.

introduction, choice architecture is usually implemented in areas with one-shot specific. Or granular behaviours, whereas eco-feedback is an ongoing regulation of behaviour, often on an aggregate level. By combining the two, however, much greater impact could be achieved. For example, choice architecture based interventions could be implemented to promote adoption of environmentally friendly products. Engaging households with eco-feedback and instilling a pro-environmental identity could then lead households to the utility's website where water retrofitting products or solar panels are promoted using choice architecture techniques. In this regard, eco-feedback does not just act as an intervention in itself but can be used as a tool for engaging people and delivering other interventions. Combining interventions may not always be successful though. For example, combining social norms and a commitment device did not promote pro-environmental behaviour as well as when each intervention was delivered separately (Terrier & Marfaing, 2015). Where combining has worked, however, is when an eco-feedback intervention was combined with more eco-feedback to tackle infrequent events. When households that were already receiving regular energy reports received additional norm based feedback regarding an upcoming energy peak load event driven by high temperatures, the two interventions generated a combined treatment effect larger than the two interventions separately (Brandon et al. 2019). In summary, the main treatment effects found in eco-feedback are modest, but according to the wider eco-feedback literature, they can have a cumulative effect, may lead to positive spillovers, and can be combined with other interventions to improve its impact. Future work should try to study the wider implications and effects of eco-feedback.

Moving beyond the average treatment effect, all the treatments proved to be heterogeneous in their effects, but not in any consistent way. Some of the analysis of heterogeneity showed increases in consumption in response to the treatment, while others showed a reduction in consumption for all. Previous studies that have looked at differences in treatment effects across different experimental context, both within a country (Allcott 2015) and between countries (Andor et al., 2020a), found that treatment effects were larger when baseline consumption was higher since there is more scope to reduce consumption, whereas low consumers would reduce by the least or may even licence themselves to increase their consumption. While this was sometimes the case, it was not consistent across all experiments. The experiment in paper 1, as well as the medium experiment in paper 2, were both implemented in the same region of the UK, and they both exhibited a similar U-shape quantiles distribution with most of the treatment effect occurring in the middle quantiles. These differed from the quantile results of the frequency experiment where almost the inverse was found, or for those in paper 3. While the results can be expected to differ between the studies in the UK and the study in the Middle East, possibly due to differences in climate, culture, and the content of feedback, it is not clear why the heterogeneity from both quantiles and CATEs differed between the two sites in the UK. Both regions have quite similar patterns of consumption and are both prone to the effects of drought. One observable difference may be the utility company that serves them. The frequency experiment is served by Anglian Water, while the medium experiment and the experiment in paper 1 are both served by South East Water. It might be the case that the messenger can impact the effectiveness of the feedback (Pornpitakpan, 2004). For example, households served by one utility company may demonstrate greater psychological reactance to the utility company's messaging causing high consumers to increase their consumption, while another utility company may have a more favourable reputation and thus are able to better motivate their high consumers. Alternatively, there may be very subtle cultural or contextual difference between the two regions of the UK that are enough to cause these differences in heterogeneity.

Relatedly, the differences in experimental context may help to explain the null result for water consumption in the Middle East. When looking at energy consumption, Andor et al. (2020a) found that, when compared to the US, countries with lower energy consumption would produce smaller effect sizes. This trend, however, does not seem to hold with regards to water consumption since the baseline water consumption in the Middle East is comparable to that of the US (>1,000 litres/day), where an eco-feedback intervention achieved a 3% reduction in water consumption (Bhanot, 2017). Even in the context of the UK where baseline consumption levels are much lower (>300 litres/day), a treatment effect of 2% was achievable. Therefore, there may be some unobserved cultural or contextual factor in the Middle East that makes water conservation even more challenging that in other contexts, and so more investigation into identifying differences in heterogeneity between sites to better understand the generalisability of findings (Meager, 2019; Bryan et al. 2021). Additionally, it may be that eco-feedback cannot be assumed to work similarly for water as it does for energy. While the two may be related and have some overlap, the conservation behaviour for water and energy can be distinct and should not be assumed to be the same. This thesis has tried to emphasise the use of quantile regressions for both having a better understanding of the treatment effects, but also to be able to identify heterogeneity. It was unexpected, yet very interesting, to often be able to identify rank invariance in the quantile results by identifying similar patterns in the CATEs. If this holds in all instances of eco-feedback interventions, quantile regressions could potentially be used to identify the pattern of heterogeneity by household consumption, even in the absence of extensive baseline data. Quantile estimates are also more robust to outliers and fat tails, as it focuses on the median rather than the mean (Koenker & Basset 1978). While some outliers were removed in paper 1 and 2 based on suspected meter read errors, the use of quantile regressions are useful in this regard since household water and energy consumption can often have extreme values.

With regards to the frequency of feedback, the frequency experiment in paper 2 demonstrated that it may be more cost-effective to deliver eco-feedback interventions less frequently than monthly without compromising treatment effects. This is somewhat supported by the experiment in paper 1, as well as the medium experiment in paper 2 which found significant and similar treatment effects even while only delivering the treatment every 6 months. Furthermore, while Allcott and Rogers (2014) found that a higher frequency of reports leads to less backsliding, and thus a longer period before the treatment effect would completely disappear once the treatment had stopped, this was not observed in paper 1 or the frequency experiment of paper 2. The studies here, of course, did not randomise households to stop receiving the reports, and so this is only an observation. Nonetheless, it appears that higher frequency does not significantly increase treatment effects, nor does it ensure longer-lasting treatment effects after the treatment has ceased. It is unclear whether these results are different from those by Allcott and Rogers (2014) because of the difference in context in which the experiments are conducted, or whether it is simply a difference in the way these interventions work on water as opposed to energy consumption. Either way, it is reasonable to assume that if the treatment effect is to be maintained, the treatment will need to persist.

Frequency, however, may be secondary to the medium at which the treatment is delivered. As shown by the non-significant effect of email in the medium experiment, and the small treatment effect found in paper 3, which was delivered by email. A more engaging medium is necessary for any effect to take place. That is not to say that email is a completely useless medium of delivering the treatment. As shown by the results of the local average treatment effects of opened emails in paper 3, if the email is engaging enough to encourage people to open them, reasonable effect sizes can be achieved. The issue is that the content of an email is often transparent and so more easily ignored. Those who did open the email and subsequently reduced their consumption may have already been motivated and were adopting feedback-seeking behaviour. One way of optimising the medium of an intervention could be to only send emails to those that are already engaged or are likely to have feedback-seeking behaviour, and to then deliver the intervention by paper for those that are not. This would reduce the overall cost of the intervention by reducing the volume of printing and postage of the reports, while maximising treatment effects where it mattered, and could be achieved by implementing an uplift model to identify these different sub-groups (Olaya et al. 2020). Alternatively, interventions could benefit from identifying ways of triggering feedback-seeking behaviour in its recipients in an attempt to manufacture engagement. Future studies could therefore test out different framing of the feedback interventions to tap into one of the four different motivators of feedback seeking behaviour (Anseel et al., 2007; Crommelinck & Anseel, 2013). As mentioned, the benefit of email-based interventions is of course that it is significantly cheaper for the utility company or government, but also that it reduces the use of resources associated with printing and posting the reports. A compromise might be to combine frequency and medium by sending paper reports every 6 months with monthly email reports in between. This ensures that the minimum treatment effect found in paper 1 and to a lesser extent, paper 2, can be achieved while creating the opportunity for higher frequency engagements that could potentially increase the overall treatment effects and promote long term habitual changes (Allcott & Rogers, 2014). A more conservative, but costlier, approach could also be to send paper reports every 3 months, with monthly email reports in between.

The results of paper 3 suggests that combining the two metrics of water and energy is not effective at reducing consumption when compared to presenting both metrics separately. This is in opposition to previous research that suggests that more abstraction or higher-level construal information is more effective at self-regulation (Fujita & Carnervale, 2012). Alternatively, the results may be more in line with other studies that have found that more granular eco-feedback (e.g., appliance level) produces larger effect sizes when compared to more aggregate eco-feedback (e.g., household level), and so delivering feedback of water and energy combined may not be granular enough to promote conservation (Andor et al. 2020b; Tiefenbeck et al. 2019). This is also supported by Feedback Intervention Theory that suggests that task focused feedback leads to better performance whereas feedback that cues the self (e.g., a combined 'eco' score) leads to less performance since it potentially threatens a person's self-esteem (Kluger & DeNisi, 1996). More research is required to better understand the role that granularity plays on the effectiveness of eco-feedback interventions. Even when presenting the energy and water separately, households seem to only reduce energy, but not water, consumption. As previously discussed, one explanation of this is that there may be something about the experimental context that makes it hard for water conservation to occur. Alternatively, households could be goal shielding and are therefore only focusing on one metric of consumption at a time. Trying to address both water and energy consumption at the same time, may best be tackled by simply focusing on one goal with a view that it will spillover into the other, either through the chain of motivation, or simply as a by-product of the water-energy nexus (e.g., Tiefenbeck et al. 2019). This could be paired with changes in frequency where households alternate between receiving water or energy reports every 3 months. It is generally assumed in the literature that eco-feedback interventions work similarly for water consumption, as they would with energy consumption, but some of the results found here in general, and in paper 3 specifically, suggests that there may be some differences. It is therefore important to further investigate how the effects of eco-feedback differ for water and energy consumption, and why presenting both at the same time does not achieve the same engagement on both. Having a better understanding of this can help in the design of interventions that try to tackle both goals simultaneously through a single intervention, leading to lower operational costs of delivering eco-feedback. Insights here could also be expanded into better understanding how to link various pro-environmental behaviours together.

One of the main limitations of all the experiments was the limited amount of demographic data that was available to better identify and interpret heterogeneity. Having more enriched data about households could potentially lead to better tailoring and targeting of feedback interventions. One that can target based on every individual's feedback-seeking self-motive, or simply one that better fits the individual's lifestyle. This can still potentially be achieved even without additional demographic data through the use of adaptive trials approach such as Sequential Multi Arm Randomised Treatment trials (Lei et al., 2012). This approach essentially randomly assigns households to different treatment groups, measures their change in behaviour after a set period, and if they have not reached a predetermined goal, they are re-randomised into a different treatment. This cycle repeats until all households are allocated to a treatment that is most effective for them. This approach can address heterogeneity without needing to necessarily know what the underlying cause of that heterogeneity is. It is therefore hoped that given the heterogeneity within and between studies that have been observed here, future versions of eco-feedback will adopt a more informed and targeted approach to maximise its impact while reducing waste.

To better understand the psychological mechanisms that drive the heterogeneity of treatment effects though, the studies would have greatly benefitted from household surveys being conducted to measure people's attitudes towards the environment, their motivation to act, their perceptions regarding the interventions themselves, as well as behavioural components to understand how they engage in water and energy consumption. This information could have contributed to developing a greater understanding of the individuals factors that moderate the efficacy of feedback interventions. For example, this type of data would've helped to explain whether households that were observed to have increased their consumption following the intervention were doing so as psychological reactance, and if so, whether that psychological reactance was related to pro-environmental attitudes. Alternatively, comparing people's pro-environmental attitudes prior to the intervention, and how it correlates with consumption, to how much change they exhibit and their subsequent attitudes, could tell us something about how feedback interacts with goal commitment (Ryan & Deci, 2000; Fishbach et al., 2006). This psychographic data would've been especially interesting given the different cultural contexts in which the studies took place, as previous research has suggested that cultural differences may have a moderating effect on social norms and sustainability (Saracevic & Schlegelmich, 2021). Most evidence of the effectiveness of eco-feedback has been from western countries (Andor et al., 2020a), and so gaining some these additional insights would've helped in developing a universal theory of how eco-feedback works.

Nonetheless, the evidence gathered in these studies do still contribute to several theories. Firstly, they show that people do change behaviours when presented with information about their peers, and this seems to hold across different cultures and context. Whether this is due to a sense of social pressure as suggested by the Focus Theory of Normative Conduct (Cialdini et al. 1990) or simply because the descriptive social norm was used as a self-regulatory goal in line with Goal Setting Theory (Locke & Latham, 2012). Second, the work here contributes towards Construal Theory by showing that more abstract or high-level construal goals are potentially less effective than lower-level concrete goals. Relatedly, as predicted by Feedback Intervention Theory, task focused feedback leads to better performance than more abstract feedback of the self (Kluger & DeNisi, 1996). And finally, the studies here highlight the importance of studying heterogeneity within a study and across different contexts to properly understand and significantly change human behaviour (Bryan et al. 2021).

People will always differ across contexts and subsequently, they will differ in their response to different interventions and designs of interventions. The generalisability of studies is not only important for understanding whether to implement interventions in other contexts but is also important for understanding how to implement interventions in other contexts. This thesis has merely scratched the surface of what is possible to explore and improve in eco-feedback but will hopefully motivate future research to explore different and new ways to design and deliver these interventions. In a world where behavioural science-based interventions are being adopted across many domains, studying ways to optimise the design and delivery of interventions should be an important avenue of research to maximise the opportunity of making a positive impact.

References

Allcott, H., & Rogers, T. (2014). "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104 (10): 3003–37.

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." Quarterly Journal of Economics 130 (3): 1117–65.

Andor, M. A., Gerster, A., Peters, J., & Schmidt, C. M. (2020a). Social Norms and Energy Conservation Beyond the US. *Journal of Environmental Economics and Management*, *103*, 102351. https://doi.org/10.1016/j.jeem.2020.102351

Andor, M., Gerster, A., & Goette, L. (2020b). Disaggregate Consumption Feedback and Energy Conservation Available at SSRN: https://ssrn.com/abstract=3638059

Anseel F, Lievens F, Levy PE. A self-motives perspective on feedback-seeking behaviour: linking organisational behaviour and social psychology research. Int J Manag Rev 2007;9 (3):211–36.

Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, *62*, 155–172. https://doi.org/10.1016/j.joep.2017.06.011

Bryan, C.J., Tipton, E. & Yeager, D.S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nat Hum Behav* **5,** 980–989. https://doi.org/10.1038/s41562-021-01143-3

Carlsson, F., Jaime, M., & Villegas, C. (2020). Behavioral spillover effects from a social information campaign. *Journal of Environmental Economics and Management*, 102325.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (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), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Crommelinck, M., & Anseel, F. (2013). Understanding and encouraging feedback-seeking behaviour: A literature review. Medical Education, 47(3), 232–241. http://doi.org/10.1111/medu.12075

DellaVigna S, Linos E. Rcts to Scale: Comprehensive Evidence from Two Nudge Units. National Bureau of Economic Research; 2020.

Evans, L., Maio, G.R., Corner, A., Hodgetts, C.J., Ahmed, S., & Hahn, U., (2012). Self-interest and pro-environmental behaviour. Nature Climate Change, 3, 122-125

Fishbach, A., Dhar, R., & Zhang, Y. (2006). Subgoals as substitutes or complements: the role of goal accessibility. Journal of Personality and Social Psychology, 91(2), 232–242. http://doi.org/10.1037/0022-3514.91.2.232

Hummel, D., & Maedche, A., (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. Journal of Behavioral and Experimental Economics, 80, 47-58.

Kluger, A. N., & DeNisi, A. (1998). Feedback Interventions: Toward the Understanding of a Double-Edged Sword. Current Directions in Psychological Science. http://doi.org/10.1111/1467-8721.ep10772989

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. A. (2012). A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology, 8,* 14.1-14.28. doi: 10.1146/annurev-clinpsy-032511-143152 PMC Journal- In process

Li, X., Zhang, D., Zhang, T., Ji, Q., & Lucey, B (2021). Awareness, energy consumption and pro-environmental choices of Chinese households. Journal of Cleaner Production, 279 (10). https://doi.org/10.1016/j.jclepro.2020.123734

Locke, E.A., & Latham, G.P. (Eds.). (2012). New Developments in Goal Setting and Task Performance (1st ed.). Routledge. https://doi.org/10.4324/9780203082744

Meager, R. (2019). Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments. *American Economic Journal. Applied Economics*, *11*(1), 57–91. http://dx.doi.org.gate3.library.lse.ac.uk/10.1257/app.20170299

Mertens, S., Herberz, M., Hahnel, U.J.J., Brosch, T., (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. Proceedings of the National Academy of Sciences, 119 (1). https://doi.org/10.1073/pnas.2107346118

Nisa, C., Belanger J.J., Schumpe, B.M., & Faller, D.G. (2019). **Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. Nature Communications, 10.**

Olaya, D., Coussement, K. & Verbeke, W. A survey and benchmarking study of multitreatment uplift modelling. *Data Min Knowl Disc* **34,** 273–308 (2020). https://doi.org/10.1007/s10618-019-00670-y

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology, 34*, 243–281.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55, 68-78.

Szaszi, B., higney, a., Charlton, A., gelman, a., Ziano, I., Aczel, B., … Tipton, E. (2022, January 19). No reason to expect large and consistent effects of nudge interventions. https://doi.org/10.31234/osf.io/mwhf3

Terrier, L., & Marfaing, B. (2015). Using social norms and commitment to promote pro-environmental behavior among hotel guests. Journal of Environmental Psychology, 44, 10-15

Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E., & Staake, T. (2019). Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nat Energy* **4,** 35–41 (2019). https://doi.org/10.1038/s41560-018-0282-1

Saracevic, S., & Schlegelmilch, B. B. (2021). The Impact of Social Norms on Pro-Environmental Behavior: A Systematic Literature Review of The Role of Culture and Self-Construal. *Sustainability*, *13*(9), 5156. MDPI AG. Retrieved from http://dx.doi.org/10.3390/su13095156