Now You Know

Using feedback from digital technology to disrupt and change habitual behaviour

Sander Hermsen
Now You Know
Using Feedback from Digital Technology
to Disrupt and Change Habitual Behaviour
The work presented in this thesis was supported by a doctoral grant from the Utrecht University of Applied Sciences.
The work presented in chapters 4, 5, and 6 of this thesis was supported by grant 057-14-010 from the Netherlands Organisation for Scientific Research (NWO).
Printing this thesis was financially supported by VU University Amsterdam and the Utrecht University of Applied Sciences.

Design and typography: Sander Hermsen
Cover illustration: Martyn F. Overwheel
Printed and bound by Ipskamp Printers bv, Enschedé

© 2019 Sander Hermsen
All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the copyright owner.
VRIJE UNIVERSITEIT

Now You Know
Using Feedback from Digital Technology
to Disrupt and Change Habitual Behaviour

ACADEMISCH PROEFSCHRIFT
ter verkrijging van de graad Doctor of Philosophy
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Sociale Wetenschappen
op woensdag 16 januari 2019 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Alexander Stephanus Hermsen
geboren te Gendringen
promotor: prof.dr. P. Kerkhof
copromotoren: dr. J.H. Frost
dr. R.J. Renes
La constance d’une habitude est en rapport avec son absurdité
– Marcel Proust, À la recherche du temps perdu. Tome 5: La Prisonnière, p. 53

Nothing so needs reforming as other people’s habits
– Mark Twain, Pudd’nhead Wilson
# Table of Contents

## Chapter 1: Introduction 8


**Introduction** 13

- Habitual Behaviour 13
- Disrupting And Changing Habitual Behaviour By Self-Monitoring And Feedback 14
- Feedback On Behaviour Through Digital Technology 15
- How Feedback Works: Mechanisms Underlying Feedback Efficacy 16
- Factors Moderating Feedback Efficacy 17
- Reviewing The Effects Of Feedback Delivered By Digital Technology 20

**Method** 21

**Results And Discussion** 22

- Methodological Issues 23
- The Effect Of Feedback Through Digital Technology On Disrupting Habitual Behaviour 24
- The Effect Of Feedback Through Digital Technology On Durable Habit Change 25
- The Effect Of Interpersonal And Intrapersonal Differences 26
- The Effect Of Feedback Technology And Properties 28
- Other Insights 31

**Conclusion** 33

**Further Research** 35

## Chapter 3: Determinants For Sustained Use Of An Activity Tracker: An Observational Study 38

**Introduction** 39

- The Effect Of Activity Tracker Usage On Physical Activity 39
- Potential Determinants Of Tracker Use 40
- Sample Size And Duration In Previous Research On Activity Trackers 42

**Methods** 42

- Study Design 42
- Activity Tracker 43
- Participants 43
- Measures 44
- Statistical Analysis 47

**Results** 49

- Fitbit Use 49
- Reasons For No Longer Using The Tracker 51
- Factors Associated With Usage 52

**Discussion** 61

- Principal Findings 61
- Reasons To Quit Tracking 61
- Relative Importance Of Predictors For Self-Tracking 62
- Limiting Factors 64

**Conclusions** 66
Chapter 1: Introduction

Undesirable habits can be very hard to change. In recent years, we have seen a growing number of digital designs that claim to provide a solution. Many of these designs (automatically) record our behaviour and give us feedback on our performance. This thesis contributes to answering the question whether feedback through digital technology is effective to change habitual behaviour. To do so, it provides a summary of a recent analysis of the current literature, and an evaluation of two existing designs for behaviour change that provide feedback on undesired habits.

In literature, habits are commonly defined as “behaviour (...) prompted automatically by situational cues, as a result of learned cue-behaviour associations” (Gardner, 2014, p. 1; Wood & Neal, 2009, p. 580). Habits help us to come to terms with the enormous complexity of everyday life, by taking away the burden of conscious deliberation from many uncritical decisions. Unfortunately, many of our habits have adverse effects on our own health and that of the planet we live on. The rigid cue-response-chain of a strong habit overrides contradictory behavioural intentions (Verplanken & Faes, 1999; Verplanken & Wood, 2006). This may lead to undesired results when habits have a satisfying short-term effect but damaging health consequences in the long run, as with snacking, a lack of physical activity, or alcohol abuse. Furthermore, since habits do not take into account current context, changed circumstances may render habits unproductive for contemporary life, even though the behaviour may have led to rewards in the past.

The major benefit of habitual behaviour is that it circumvents active consideration of the current context, but this also makes it very hard to change habits using interventions aimed at controlled processing, e.g. through persuasive messages related to the consequences of behaviour (Verplanken & Wood, 2006) or changing behavioural intentions (Sheeran, 2002). A more successful way to change undesired habits is to bring
habitual behaviour and its context to (conscious) awareness. Self-monitoring, the procedure by which individuals record the occurrences of their own target behaviours (Nelson & Hayes, 1981), enables perception of our own behaviour and adaptation to the current context. This leads to a decrease in unwanted behaviour (Quinn, Pascoe, Wood, & Neal, 2010). Unfortunately, self-monitoring is difficult for even the most motivated individual (Wilson, 2002). There is often a discrepancy between self-reported and actual performance in health behaviours such as calorie intake and physical activity (Lichtman et al., 1992).

Accurate self-monitoring is greatly improved by personalized information from external sources (Kim et al., 2013; Li, Dey, & Forlizzi, 2010). The advent of mobile and interactive media has given us an unsurpassed opportunity to support people in self-monitoring, by providing them with tailored feedback, or “actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance” (Kluger & DeNisi, 1996), on their behaviour. Digital technology can offer constant, real-time updates, powered by sensitive measurement devices, often worn on the body. Besides data generation, digital technology can offer habit-disrupting cues such as light and sound signals, buzzes, and push messages. Digital technology is not only useful to present users with evaluations of past behaviour (“reflection-on-action”, Schön, 1983); because of the ubiquity of wearables and mobile devices, feedback from digital technology offers an unprecedented opportunity for “reflection-in-action” (ibidem): the analysis of behaviour as it occurs. This could greatly increase people’s efficacy in self-managing healthy behavioural change.

The rapid rise of the technological possibilities has been matched by a similar rise in the number of designs on the market that make use of these possibilities. Wearable activity trackers (cf. Kooiman et al., 2015) give us feedback on whether we walk enough; sleep monitors monitor the quality of our sleep (e.g. Ogihara & Eshita, 2015); smart devices track our eating habits (Zandian, Ioakimidis, Bergh, Brodin, & Södersten, 2009), an app can warn us about situations in which we are likely to smoke a cigarette (e.g. Naughton et al., 2016) and a growing number of devices tell us (and others) what emotions we experience in cases where we are unable to do so ourselves (e.g. Van Dijk, 2017). This increased attention in health design practice is closely followed by a growing body of literature in design research and human-computer interaction research in the past decades (Darby, 2013; Fischer, 2008; Froehlich, Findlater, & Landay, 2010; Gouveia, Karapanos, & Hassenzalh, 2015; Hänsel, Wilde, Haddadi, & Alomainy, 2015). By far the largest part of this literature examines the different channels, modalities, and other properties of the digital feedback: how to optimally design the feedback technology.

Considering all this attention, it may come as a surprise that there has been relatively little research into whether all this feedback on health behaviour is as effective as we implicitly presume. After all, the rise in designs and research based on these
designs may very well be a case of technocratic solutionism (Morozov, 2013): we have sensors and actuators, especially in smartphones, and we have wearables. Now that we have been provided with these hammers, we suddenly see nails everywhere. But are these really nails?

In this thesis, I investigated whether feedback through digital technology is an effective way to support people in changing their undesired, unhealthy habitual behaviour. Theory supports the hypothesis that it does; with Control Theory (Carver & Scheier, 1985) delivering the best explanation: reflective behaviour change resembles a thermostat. When looking to change their behaviour, people compare their performance to a behavioural goal. When a discrepancy between goal and performance is noted – given enough motivation, opportunity, and the right abilities – people will attempt to reduce this discrepancy. This process depends on conscious scrutiny of behaviour and its effects. Knowing that habitual behaviours are mostly automatic, and thereby outside of conscious scrutiny, the strength of feedback lies in delivering exactly that cue that is needed to make automatic behaviour available for conscious deliberation. Feedback may also increase motivation to change the target behaviour (Northcraft, Schmidt, & Ashford, 2011): feedback places the target behaviour higher on a hypothetical list of priorities. When given feedback on the number of steps we take, we may prioritise walking over other modes of transportation or other physical activity choices, because feedback diverts our attention towards this behaviour.

The question is, of course, whether practice follows theory. To find out, we examined the available evidence from literature, to evaluate whether current literature provides an answer to the following questions:

Is feedback through digital technology an effective way to change habitual behaviour?

Is feedback through digital technology effective for each user in every context, or are there intrapersonal (e.g. character traits, psychological states such as motivation) or interpersonal (contextual or systemic) moderators? What feedback properties are most effective in different circumstances?

In Chapter 2 of this thesis, I describe the results of our literature review.

To provide further answers to the above questions, we then evaluated two existing designs for behavioural change. Inclusion criteria for the designs were: a) the design addresses habitual behaviour, b) the design uses feedback on behavioural performance as its (primary) behaviour change technique, c) the design can be tested in real-life conditions (beyond the lab). To obtain valid results, we only included participants who could reasonably be expected to be motivated to change their behaviour, for instance because they chose to purchase or download the design of their own accord.

Firstly, we looked at the sustained use of an activity tracker to investigate which of a broad range of intra-personal and inter-personal determinants influenced longer
use of feedback technology. The results of this study are in Chapter 3 of this thesis. Secondly, we evaluated the acceptability and user experience (Chapter 4), ability to disrupt detrimental habitual behaviour in a single setting (Chapter 5) and ability to durably change detrimental behaviour (Chapter 6) of a digital device that provides feedback on eating rate.

In Chapter 7, I draw conclusions from my research, I discuss the potential, but also the limitations and unwanted consequences of feedback from digital technology on habits and propose challenges and further research for both academia and practice.
Chapter 2: Using feedback through digital technology to disrupt and change habitual behaviour: A critical review of current literature

Abstract

Habitual behaviour is often hard to change because of a lack of self-monitoring skills. Digital technologies offer an unprecedented chance to facilitate self-monitoring by delivering feedback on undesired habitual behaviour. This review analysed the results of 72 studies in which feedback from digital technology attempted to disrupt and change undesired habits. A vast majority of these studies found that feedback through digital technology is an effective way to disrupt habits, regardless of target behaviour or feedback technology used.

Unfortunately, methodological issues limit our confidence in the findings of all but 14 of the 50 studies with quantitative measurements in this review. Furthermore, only 4 studies tested for (and only 3 of those 4 found) sustained habit change, and it remains unclear how feedback from digital technology is moderated by receiver states and traits, as well as feedback characteristics such as feedback sign, comparison, tailoring, modality, frequency, timing and duration.

Introduction

A variety of digital solutions to help us change detrimental or outdated habitual behaviour have arrived on the market. These so-called quantified self-solutions, also known as persuasive technologies, aim to alter ingrained habits by presenting people with behavioural feedback through mobile and interactive devices and applications. These technologies can help individuals improve their health and the environment by increasing awareness and improving the self-regulation of behaviour, something that does not come easily to us. Opportunities to incorporate such technologies in daily life have risen dramatically in recent years. In many nations, a great share of the general populace owns a smartphone or other kind of smart device and seems willing to use technology to change unwanted behaviours. For instance, more than 69% of US citizens track at least one health behaviour, with 14% using a specialized tracker (Fox & Duggan, 2013). Manufacturers are jumping on this bandwagon, offering new ways to measure behaviour, e.g. through Apple’s Research Kit (Moynihan, 2015).

Few of these quantified self-products have been tested in controlled circumstances (Cowan, Bowers, Beale, & Pinder, 2013). Moreover, most solutions lack scientific evidence, with positive anecdotal reports from practice comprising the basis of our understanding (Cowan et al., 2013; Schoffman, Turner-McGrievy, Jones, & Wilcox, 2013). As yet, the potential of digital technology to disrupt and possibly even change habits through feedback on habitual behaviours remains unclear.

This review addresses this gap in the literature by presenting a review of existing studies on the use of feedback generated by digital technology to disrupt and change automatic, habitual behaviours. This review adds to the current debate by providing an overview of existing evidence, accentuating and addressing gaps in current knowledge and laying an evidentiary foundation for digital technology solutions aimed at habit change.

To do so, we first assess the drawbacks of habitual behaviour and the strategies that may be applied to disrupt undesired habits. Second, we then discuss the role of self-monitoring in habit disruption and the role feedback from external sources can play in self-monitoring. In the third section, we look at known influences of feedback efficacy, and consider whether insights into the effect of feedback on habitual behaviour in general are valid when applied to feedback delivered through digital technology. Finally, we review findings on the use of digital technology that utilizes feedback and suggest avenues for future research.

Habitual behaviour
In everyday life, habits, commonly defined as “behaviour (...) prompted automatically by situational cues, as a result of learned cue-behaviour associations”
Chapter 2

(Gardner, 2014 p. 1; Wood & Neal, 2009 p. 580), help us to come to terms with the enormous complexity of everyday life. However, some of the biggest threats to personal and planetary wellbeing are direct consequences of our habitual behaviour. The cue-response-chain of a strong habit is a rigid structure, which overrides contradictory behavioural intentions (Verplanken & Faes, 1999; Verplanken & Wood, 2006). This may lead to undesired results when cue-response-pairs have a satisfying short-term effect but lead to damaging consequences in the long run, as with snacking or alcohol abuse. Furthermore, since habits do not take into account current context, changed circumstances may render habits unproductive for contemporary life, even though the behaviour may have led to rewards in the past.

Because habitual behaviour circumvents active consideration of the current context, it is hard to change habits using interventions aimed at controlled processing, e.g. through persuasive messages (Jager, 2003; Verplanken & Wood, 2006). One powerful strategy to disrupt habits is therefore to change the circumstances so that habit cueing does not occur (Verplanken & Wood, 2006) or to alter the external cues that lead to habit execution (e.g. in Aarts & Dijksterhuis, 2003). However, these strategies have practical difficulties, since manipulating or avoiding cues is often impossible (Quinn et al., 2010) and not always seen as ethical, because receivers may not always consciously notice the manipulations, which places their consequences outside the reach of conscious scrutiny (Verbeek, 2006).

Disrupting and changing habitual behaviour by self-monitoring and feedback
The automaticity of habitual behaviour means that execution is often at least partially unconscious and may start without conscious intent. Therefore, one way to disrupt undesired habits is to bring habitual behaviour and its context to (conscious) awareness. Self-monitoring, the procedure by which individuals record the occurrences of their own target behaviours (Nelson & Hayes, 1981), enables perception of our own behaviour and adaption to the current context. Thus, self-monitoring leads to decreases in unwanted behaviour (Quinn et al., 2010).

Unfortunately, self-monitoring is difficult for even the most motivated individual (Wilson, 2002). For example, there is often a discrepancy between self-reported and actual performance, as shown in diverse behaviours such as calorie intake (Lichtman et al., 1992), weight and BMI - especially in overweight participants (Pursey, Burrows, Stanwell, & Collins, 2014), the amount of exercise (Lichtman et al., 1992), actual versus perceived water use (Hamilton, 1985; Millock & Nauges, 2010); and even the reporting of relatively stable personal data such as height (Pursey et al., 2014).

Accurate self-monitoring is greatly improved by personalized information from external sources (Kim et al., 2013; Li, Dey, Forlizzi, Höök, & Medynskiy, 2011). The intentional delivery of such information about performance or behaviour (or about the impact of one’s performance or behaviour) in order to facilitate behaviour change
Using feedback through digital technology to disrupt and change habitual behaviour

is commonly referred to as feedback (Van Velsor, Leslie, & Fleenor, 1997, p. 36). In this review, we adopt the definition of feedback offered by Kluger and Denisi (1996), seeing feedback as “actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance”.

The beneficial effect of feedback on performance has been established in a range of fields. In education, the role of feedback is especially well established. Hattie and Timperley (2007) performed a synthesis of meta-analyses of feedback in educational contexts and reported an average effect size of 0.79 for feedback interventions, almost twice the average effect size of general educational interventions (0.40). This implies that feedback interventions in general are not only capable of disrupting undesirable habits but can also play a significant role in changing those behaviours. Similarly, feedback has been shown to be effective in an increasing range of controlled studies regarding both health (Gardner, Whittington, McAteer, Eccles, & Michie, 2010) and sustainability (Darby, 2013; Fischer, 2008; Froehlich et al., 2010).

Feedback on behaviour through digital technology

Direct, instant feedback used to be difficult to deliver regularly on a large scale. The delivery of feedback was restricted to either distant, impersonal media such as utility bills and letters, or cost-intensive face-to-face communication with trained personnel. The advent of mobile and interactive media has changed that. In recent years, technological developments have enabled a surge of behaviour-changing interventions. A range of mobile apps, wearable devices, web-based platforms and in-home displays give us feedback on our behaviour and monitor behaviour that previously remained hidden. There are apps and wristbands to support us in physical exercise, applications for weight loss, in-home displays to encourage us to use less energy, etcetera. Already, more than half of smartphone users gather health-related data with their phone, one in five has installed at least one health-behaviour related app (Fox & Duggan, 2013) and one in ten Americans owns some sort of automatic activity tracker (Ledger & McCaffrey, 2014). Similarly, many European countries aim to achieve smart energy meter installation in every home by 2020 (Faruqui, Harris, & Hledik, 2010).

Digital technology can offer constant, real-time updates on our progress, powered by sensitive measuring devices, often worn on the body. The widespread use of sensing systems means that automatically generated data about the undesired behaviours can be made available, without the need for possibly problematic self-reporting. Monitoring devices can be used for a range of data-gathering causes including health statistics like heart rate, blood pressure, and blood sugar (Verplanken & Wood, 2006).

---

1 This definition excludes non-task-related feedback (“he just does not like you”), and intrinsic, task-generated feedback (e.g. getting coffee from a machine and seeing that your cup is full), whilst including feedback on how a task is performed (e.g. “you kicked the ball with the tips of your toes; you should have used the instep” in football training).
and environmentally important data on energy use (Froehlich et al., 2010; Verplanken & Wood, 2006).

Besides data generation, digital technology can offer habit-disrupting cues such as light signals, buzzes, beeps, and push messages. Digital technology is not only useful to present users with evaluations of past behaviour (“reflection-on-action”); because of the ubiquity of mobile and handheld devices, digital technology offers an unprecedented opportunity for “reflection-in-action” (Schön, 1983), the analysis of behaviour as it occurs.

The availability of interactive displays provides ample opportunity for new types of feedback. A power socket may be enhanced to report energy use (Heller & Borchers, 2011), a shower head can give us feedback on water use or shower time (Andler, Woolf, & Wilson, 2013), or a power cable can move around as if in agony if connected devices are left in stand-by mode (Laschke, Hassenzahl, & Diefenbach, 2011).

Digital technology has a number of distinct advantages over human persuaders. Devices can be (irritatingly) persistent, guarantee greater anonymity and have access to areas where people are not welcome (e.g. the bedroom or bathroom) or unable to go (e.g. inside clothing or household appliances). Moreover, digital technology is relatively easy to replicate, distribute and tailor to specific needs (Fogg, 2003). However, there are some disadvantages: digital technology is a lot easier to ignore or shut down than messages delivered by human persuaders. Furthermore, digital technology solutions are easily forgotten, lost or otherwise misplaced. For example, over half of those that have owned a wearable fitness tracker no longer use it, and a third of the users quits use in the first six months after purchase (Ledger & McCaffrey, 2014). Yet, in providing automatically delivered feedback for habit change, the benefits of digital technology may very well outweigh the disadvantages.

How feedback works: Mechanisms underlying feedback efficacy

Control theory provides insight into the mechanisms underlying the effect of feedback (Carver & Scheier, 1985). According to control theory, reflective behaviour change processes are reminiscent of a thermostat. When looking to change their behaviour, people compare their performance to a behavioural goal. When a discrepancy is noted, given enough motivation, opportunity, and the right abilities, people will attempt to reduce this discrepancy. The efficacy of this regulatory cycle is moderated by three executive function skills (cf. Hofmann, Schmeichel, & Baddeley, 2012): keeping a goal salient in working memory or bringing the goal back to working memory when needed; the ability to inhibit undesired automatic responses; and the ability to switch between tasks or mental sets.

Feedback supports reflection by increasing knowledge and awareness of behaviours and their impacts. Many behaviours are of such automaticity, that their performance is at least partly subconscious. Knowing that and when a habit occurs opens up
Using feedback through digital technology to disrupt and change habitual behaviour

possibilities for behaviour change. Feedback also enables us to compare the consequences of our behaviour to our current goals and adapt when the behaviour does not fit the context. Furthermore, it also serves to increase general self-awareness, which in turn increases our capabilities to inhibit undesired behaviours (Alberts, Martijn, & de Vries, 2011).

Feedback also has motivational consequences. We are driven by motivations to approach experiences that are expected to be pleasurable, and avoid unpleasant experiences (Elliot & Covington, 2001; Higgins, 1997). Both the negative emotions caused by an observed increasing discrepancy between goals and performance, and the positive emotions caused by a decreasing discrepancy, can increase our motivation to reach our goals (Carver & Scheier, 2011; Deci, Koestner, & Ryan, 1999). Furthermore, among competing behaviours, those supported by feedback are given priority over those without feedback (Northcraft et al., 2011).

Factors moderating feedback efficacy
In a meta-analysis of 607 studies, Kluger and DeNisi (1996) found that, generally speaking, two thirds of all feedback interventions increased performance. However, the remaining third of the interventions had an opposite, detrimental effect on performance. Importantly, this means that even though we can expect a habit-disrupting effect from well-designed feedback interventions, this does not automatically signify that the feedback intervention will lead to change in the desired direction. Furthermore this suggests that an interplay of receiver states and traits on the one hand, and feedback properties such as content (e.g. sign, comparison and level of detail), timing, modality, frequency, duration, and presentation on the other, influence feedback effectiveness (Fischer, 2008). The moderating effects of both receiver traits and states and feedback properties will be discussed below.

Interpersonal and intra-personal differences
Feedback efficacy is moderated by all kinds of characteristics of the feedback receiver, in an interplay of stable and more dynamic factors. A great deal of the expected moderators is stable and relatively uncontrollable, such as socio-economic status (Maitland, Chalmers, & Siek, 2009) and gender (e.g. Guadagno & Cialdini, 2007; Ho et al., 2013).

In any self-control mechanism, executive control capabilities play an important role, such as the capacity for self-regulation. Differences in personality and context determine the degree to which an individual is capable of exercising such control (Baumeister & Heatherton, 1996; Braverman, 2008; Kuhl, 1985). In addition, self-regulating capacity is in finite supply (Baumeister, Bratslavsky, Muraven, & Tice, 1998).

Feedback efficacy is also influenced by relatively fleeting states such as high initial engagement with the target goal, strong motivation or a high perceived self-efficacy (Bandura, 1994). Self-regulation processes are cyclical in nature (Bandura, 1994;
Zimmerman, 1998). This indicates that high initial motivation leads to a greater feedback effect, which in turn leads to increased motivation (e.g., Geister, Konradt, & Hertel, 2006). Similar cyclical effects can be found for self-regulatory skills and perceived self-efficacy (e.g., Donovan & Hafsteinsson, 2006; Multon, Brown, & Lent, 1991).

To date, there is little or no evidence on whether these intra- and interpersonal factors that are generally known to influence feedback efficacy, such as motivation and perceived self-efficacy towards the goal, self-regulatory capabilities, and demographic and socio-economic factors, have different effects on the efficacy of feedback when it is delivered through digital technology. Since the latter is generally delivered in an individual context and not within the social setting of interpersonal feedback, the effect of feedback through digital technology might rely on capabilities and motivation of the receiver more than with interpersonal feedback.

Feedback properties
Paying attention to carefully crafting the timing, delivery, and content of the feedback can enhance the effectiveness of feedback interventions. In an extensive review of feedback on household energy use, Fischer (2008) indicates that high frequency feedback delivered over a long period by computerized and interactive tools provides an advantage in feedback effectiveness. There are a number of feedback properties that may affect effectiveness, including technology, content, timing, modality, duration, frequency, and presentation and user experience. Generally, the largest effects can be expected from detailed, positively framed, concurrent feedback (‘reflection-in-action’), delivered continuously or on-demand through more than one modality, during a long period.

Technology. Feedback can be delivered through many different technological channels, ranging from websites and smartphone apps to wearables and in-home displays. The possibility to deliver well-designed and automatically tailored, in-action, frequently delivered feedback over a long period of time is one of the perceived strengths of digital, interactive technology. Because behaviour often is measured directly, a direct response is possible, and the all-pervasive use of smartphones and other technologies means instant delivery on a large scale is relatively easy.

Each form of the technology has its advantages and disadvantages as a source of feedback. For example SMS text messages, a well-researched and generally considered effective means of feedback delivery (Hall, Cole-Lewis, & Bernhardt, 2015), are difficult to deliver at the very moment the behaviour occurs because of time lag. This delay can severely disrupt performance, which may in some cases have negative consequences on behavioural fluency (Bittner & Zondervan, 2015). Furthermore, text messages can only deliver content of limited length (usually about 160 characters). On the other end of the spectrum, wearable activity trackers can do real time tracking of
Using feedback through digital technology to disrupt and change habitual behaviour

behavioural data, and are capable of on-demand or continuous delivery over a range of sensory channels without limits to the richness of the data (Yang & Hsu, 2010).

Content. Tailoring content to fit receiver characteristics can be expected to affect feedback effectiveness. Ample evidence from the literature shows that tailoring message content to meet recipient motivation, traits, abilities and preferences increases the effectiveness of such messages (Ivers et al., 2012; Kaptein, De Ruyter, Markopoulos, & Aarts, 2012; Noar, Grant Harrington, Van Stee, & Shemanski Aldrich, 2011; e.g. Noar, Benac, & Harris, 2007). Such tailoring may encompass utilizing negative, positive or neutral feedback (i.e. feedback sign); offering social, historical or normative comparisons (or no comparison at all); and increasing or decreasing level of detail.

Timing. There has been substantial research on the effect of feedback timing on learning (Hattie & Timperley, 2007). Specifically, reflection-in-action can be expected to be more effective than reflection-on-action. For instance, in electricity use, direct, short delay feedback on energy usage generally leads to a 5–15% reduction in consumption, and indirect, long delay feedback leads to a reduction of 0–10% (Darby, 2013).

Modality. Selecting optimal delivery through visual, auditive, or tactile channels, or a combination of channels, increases feedback effectiveness (Braverman, 2008; Hoggan, Crossan, Brewster, & Kaaresoja, 2009; Warnock, McGee-Lennon, & Brewster, 2011). An optimal modality choice depends on the possibility of disruption and the need for detail. The visual mode is more disruptive than the auditory, which is in turn more disruptive than tactile feedback. Similarly, visual feedback can contain more detailed information than auditory, which in turn has more capacity for detail than tactile feedback.

Frequency and duration. Frequency and duration of the feedback intervention also influence feedback effectiveness. In general, the more frequent the feedback is delivered, over a longer period of time, the more the intervention will contribute to behaviour change. The benefits of more frequent feedback are limited by cognitive capacity: as long as the frequency of the feedback does not overwhelm an individual’s cognitive resources, more feedback is better (Lam, DeRue, Karam, & Hollenbeck, 2011). Current technological developments, especially those that concern use of mobile and interactive platforms, make it possible to circumvent these limitations and easily deliver much more frequent or even continuous feedback, with infinite durations. In theory, this should increase feedback effectiveness.

Presentation and user experience. Research in web design (Tuch, Presslaber, Stöcklin, Opwis, & Bargas-Avila, 2012), typography (Larson, Hazlett, Chaparro, & Picard, 2005) and usability (Tractinsky, Katz, & Ikar, 2000) suggests that visual design aspects and aesthetics determine the attitude towards a design as well as the perceived ease of use (but not actual use). Consequently, users will feel more beneficial towards an
Chapter 2

aesthetically pleasing intervention and will be more inclined to persevere in using it. Moreover, a clear design might aid in emphasizing important information, personalizing the feedback and improving the fluency of feedback. However, the design and presentation of the feedback and technology must also fit participants’ goals. For example, research on the design of glucometers suggests that the desired look and feel depends on context; users favor a more “medical” appearance when passing through customs on transatlantic flights and inconspicuous or sporty looks in day to day life (O’Kane, Rogers, & Blandford, 2015).

Reviewing the effects of feedback delivered by digital technology
Feedback through digital, interactive technology can have two beneficial effects on habitual behaviour. Firstly, it can disrupt the automatic execution of the habitual behaviour, making it available for conscious scrutiny. Secondly, feedback can lead to durable behaviour change. Given the extensive evidence for the beneficial effect of feedback on habitual behaviour change in general (Brug, Glanz, Van Assema, Kok, & van Breukelen, 1998; Fischer, 2008; Hattie & Timperley, 2007; Ivers et al., 2012; Kluger & DeNisi, 1996), and the aforementioned benefits of digital technology over more traditional forms of feedback delivery, one assumption in this work is that feedback delivered by digital technology is at least as effective as ‘regular’ feedback in disrupting undesired habits. Furthermore, based on literature on feedback on habitual behaviour in general, feedback delivered through a well-chosen digital technology appears well suited to increase the chances of durable, lasting behaviour change.

However, the fact that feedback through digital technology is delivered without the intervention of a human source might influence its effect, e.g. because of the lack of social pressure. Similarly, the effects of receiver moderators such as motivation and perceived self-efficacy are likely, but not certain, to be similar to those reported for feedback in general (the more motivation or the higher the perceived self-efficacy, the more effect of feedback can be expected).

The current review provides an overview of recent original studies that look into the effect of feedback through digital technology on undesired habitual behaviours. This review provides an analysis of the efficacy of such feedback to both disrupt and durably change habitual behaviour. Furthermore, the review evaluates the effects of interpersonal and intra-personal differences; technology choice; and feedback properties: technology, content, timing, modality, duration, frequency, and presentation and user experience, on feedback efficacy.
Using feedback through digital technology to disrupt and change habitual behaviour

Method

A combined search of the databases PubMed, PsychInfo, EMBASE and Web of Science was performed with the following set of search terms: (habit* OR habitual behaviour) AND (persuasion OR behaviour change OR habit disruption) AND (feedback OR self-monitoring) AND (persuasive design OR persuasive technology OR digital technology). This search resulted in 993 results. The ACM Digital Library and the IEEE Xplore Digital Library were searched, using the search terms “feedback AND persuasive AND habit”. This search yielded 416 results from ACM/DL and 233 results form IEEE/Xplore; these results included peer-reviewed journal papers as well as conference proceedings.

Abstracts from both result sets were checked for relevance. From these, 101 publications with relevant and ambiguous abstracts were retained. Papers cited in included articles were checked for eligibility. Ancestry searches were performed on the included articles through Google Scholar, to retrieve more recent articles building upon the original work. From these searches, a further 35 primary publications were included. This resulted in a set of 136 primary sources.

From this set, 69 original papers matched the following inclusion criteria:
• The research has the primary purpose of changing habitual behaviour, either increasing or decreasing the behaviour or stopping the behaviour altogether. Habit is operationalized as recurring behaviours with some degree of automaticity (Wood & Neal, 2009).
• Digital technology has to be used as the primary means of achieving behaviour change. The digital technology must use a tailored feedback mechanism delivered by (an) external agent(s) to provide information regarding task performance.
• The research must encompass some form of analysis of the effect of the intervention on the targeted behaviour, be it qualitative or quantitative.
• Because of rapid developments in the field of digital technology, only papers from the last decade (2004 and later) were included.

All analyzed papers are included in the reference list and marked with an asterisk (*). One included paper reported three relevant studies (Nakajima & Lehdonvirta, 2013) and two papers reported two relevant studies (Connelly, Faber, Rogers, Siek, & Toscos, 2006; Stienstra, Alonso, Wensveen, & Kuenen, 2012), all of which were separately scored. This resulted in a final set of 72 studies.

The broad range of dependent variables, feedback intervention technologies, and research methods applied in the included studies made it impossible to conduct a meta-analysis of results in such a way that it would produce reliable and valid insights (Borenstein, Hedges, Higgins, & Rothstein, 2009; Quintana, 2015). Consequently, a systematic review with a descriptive analysis (Garg, Hackam, & Tonelli, 2008) of the literature was performed. Even though, when compared to a meta-analysis, a
systematic literature review has more limited possibilities to derive general conclusions, this approach is able to shed light on the general direction of effects, as well as identify gaps in the literature (ibidem). Furthermore, conducting a systematic literature review enables us to incorporate results from qualitative studies, which would not be possible in a meta-analysis.

We thematically classified target behaviours of the intervention, feedback technology, feedback characteristics (content (feedback sign, comparison, and level of tailoring), timing, modality, frequency, duration, data source), and the availability of visual examples of the design and provided feedback. For each intervention, number of participants, independent and variables, analysis method, results, and possible methodological concerns were scored.

The included studies covered a range of dependent variables, varying from energy consumption to motor skills and physical activity. A list of the occurrence of each category of dependent variable is included in Table 2.1. A full list of included studies, including target behaviours, feedback content, characteristics, dependent and independent variables and measurement methods is available as an online appendix (see list of supplementary materials, Appendix 1).

<table>
<thead>
<tr>
<th>Occurrences</th>
<th>Dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>energy and water consumption</td>
</tr>
<tr>
<td>11</td>
<td>motor skills (speed skating, posture, violin playing, tooth brushing)</td>
</tr>
<tr>
<td>10</td>
<td>healthy eating and weight loss</td>
</tr>
<tr>
<td>9</td>
<td>physical activity</td>
</tr>
<tr>
<td>6</td>
<td>driving</td>
</tr>
<tr>
<td>3</td>
<td>general wellbeing</td>
</tr>
<tr>
<td>3</td>
<td>waste reduction</td>
</tr>
<tr>
<td>2</td>
<td>break-taking and resuming work</td>
</tr>
<tr>
<td>9</td>
<td>other (social feedback, bookshelf ordering, IQ training, printing behaviour, medication adherence, overfilling water cookers, transport mode choice)</td>
</tr>
</tbody>
</table>

### Results and Discussion

In this section, we first discuss the consequences of the diverse methodological approaches, followed by an analysis of review results ordered by theme – general effects of feedback on disrupting and changing habitual behaviour, the effect of receiver characteristics, and the effects of different feedback technologies and characteristics. Finally, we discuss a few insights that transpired from qualitative results that were not based on a pre-posed hypothesis.
Methodological issues
The broadness of the range of studies included in this review is reflected in the different methodological approaches used. Of the 72 studies included in this review, three studies took place under controlled (laboratory) circumstances, 20 were field studies (7 of which were set up as a randomised controlled trial), and 49 studies tested a prototype or design. With regard to methods of analysis, 21 studies used qualitative analysis, mostly user experience studies describing interactions with designed prototypes. 50 studies utilized some form of quantified measurement and analysis, in 15 cases together with qualitative measures. In one study, data gathering and analysis were described so poorly, that it remained unclear which research methodology was used.

Each form of research design and method of analysis has its own unique merits to the generation of knowledge. However, in every research design, reliability and validity should be well thought-through, to prevent experimental artifacts such as the Hawthorne effect – mere observation enhancing performance (cf. McCarney et al., 2007) –, demand characteristics – participants’ interpretation of what is expected of them (Orne, 2009), or unforeseen events influencing performance – such as seasonal influences on energy use that may eclipse the effect of a feedback intervention. In general, quantitative studies that include (active) control groups, pre- and post-test measures, and use a fitting statistical test with ample power (Maxwell & Delaney, 2004, p. 56–59) are better suited for this. In qualitative study designs, a well-structured data collection and analysis strategy is necessary to reduce the chance of cherry-picking precisely those results that fit the hypothesis (Patton, 1990).

Most of the included quantitative studies did not meet these criteria. 33 of 50 quantitative studies did not report a strategy of dealing with experimental artefacts such as demand characteristics or unforeseen external moderators. Of the 50 quantitative studies, 30 studies were analysed using statistical testing, yet only 8 out of these 30 studies showed sufficient statistical power for the sort of analysis performed. This is important, since low statistical power implies a large chance of type I and II errors (Cohen, 1992). Furthermore, low statistical power combined with a significant result dramatically increases the chance of an overestimation of intervention effects (Gelman & Carlin, 2014). In total, only 14 out of 50 studies with some sort of quantitative measurements had sufficient statistical power plus an experimental design that would prevent the occurrence of the most common experimental artefacts.

The 21 qualitative studies included in this review were all of sufficient rigor to avoid cherry picking in results. Most studies used a form of structured interviewing as data collection method and reported some sort of systematic appraisal of the results. No qualitative studies had obvious methodological shortcomings.
We focus our analysis on those studies that meet all criteria mentioned above, both utilizing qualitative and quantitative methods. Subsequently, we will mention descriptive results from studies that did not meet all these criteria, with a corresponding caveat.

The effect of feedback through digital technology on disrupting habitual behaviour
The effect of feedback through digital technology on disrupting habitual behaviour is generally confirmed by our analysis. Of the 72 studies included in this analysis, 59 studies show a beneficial effect of feedback on disrupting habitual behaviour. 13 of the 14 studies with well-set up experimental designs and ample statistical power report significant results. 25 studies show a beneficial effect based on qualitative measurements, including observation reports, interviews and other user experience measures. Furthermore, 32 of the 37 (quantitative) studies with methodological issues studies report descriptive data that point in the direction of hypothesis. Of all studies that report a beneficial effect, five studies found this effect to be partial, i.e. not occurring in every expected condition.

Thirteen of fourteen well set-up experimental studies prove the beneficial effect of feedback through digital technology on a broad range of habitual behaviours. Feedback increased fruit consumption (Bech-Larsen & Grønhøj, 2013), safer driving behaviour (Donmez, Boyle, & Lee, 2008; Maltz & Shinar, 2007), motor learning (Lieberman & Breazeal, 2007) and posture training (Epstein, Colford, Epstein, Loye, & Walsh, 2012), lowering eating rate (Ford et al., 2009), increasing physical activity (Hurling et al., 2007; Schulz et al., 2014), weight loss (Pellegrini et al., 2012; Schulz et al., 2014), limiting computer use (Van Dantzig, Geleijnse, & Van Halteren, 2013), shower use (Willis, Stewarta, Panuwatwanich, Jones, & Kyriakides, 2010), and electricity consumption (Jain, Taylor, & Peschiera, 2012; Vassileva, Odlare, Wallin, & Dahlquist, 2012; G. Wood & Newborough, 2003).

One well-designed quantitative study reported a null effect. The lack of effect in this study, in which participants could volunteer to join a home energy reduction intervention (Alahmad, Wheeler, Schwer, Eiden, & Brumbaugh, 2012), could be ascribed to ceiling effect caused by participant self-selection, such that only highly motivated participants that already performed many energy-saving behaviours took part. This could prove a limitation of the efficacy of feedback interventions: when participants are already performing the behaviour in some way, there is a limit to habit change coming from feedback.

Eight qualitative studies reported no effects or even a contrary effect of feedback on behaviour change. One study on waste disposal (Comber & Thieme, 2013) and a study on electricity usage (Hargreaves, Nye, & Burgess, 2010) found that although no behaviour change was registered, knowledge about which behaviours were desirable and which less so did increase. In two studies, participants
did not understand the manipulation (Gyllensward, Gustafsson, & Bang, 2006; S. Kim, Kientz, Patel, & Abowd, 2008). One further study (Nakajima & Lehdonvirta, 2013) on utilizing feedback to encourage a certain ordering of books on a bookshelf, led participants to play around with the installation, with inverse effects. Inverse effects were also found in a study on taking breaks at work, where participants used social activity feedback to avoid colleagues or to find empty rooms for meetings (Kirkham et al., 2013). This, too, may be a limitation of feedback: receivers may not perceive the feedback as a cue towards the target behaviour. Studies by Katzef et al. (2013) on energy use in the office, and Strengers (2011) on energy and water consumption show how feedback may not per se lead to behaviour change, but may in the latter case also cause post-hoc rationalizations of the undesired behaviour.

Finally, four quantitative studies found null results; however, all four studies (Cowan, Bowers, Beale, & Pinder, 2013; Pereira, Quintal, Nunes, & Bergés, 2012; Quintal, Pereira, & Nunes, 2012; Rodgers & Bartram, 2011) suffered from a lack of statistical power, so their null finding may very well be due to small sample sizes, since descriptive results in all studies do show a small positive effect of the reported interventions.

Where possible, we calculated effect sizes of quantified measurement methods for comparison (Table 2.2). 28 studies either reported effect sizes or presented their data in such a way that effect sizes could be calculated. Even though the broad range of dependent and independent variables used in the reviewed studies make direct comparison in the form of a meta-analysis unfeasible, an overview of effect sizes listed could in theory serve as an indication of effect sizes to be expected in feedback interventions on habitual behaviour.

Because of the methodological issues in the greater part of these studies, the reported effect sizes should be used with extreme caution. Low statistical power, especially, increases the chance of inflated effect sizes (Gelman & Carlin, 2014), which would give at least a partial explanation of the size of the effects found in many studies in this review.

The effect of feedback through digital technology on durable habit change
The durability of the hypothesized effect was tested in only four of the 72 studies, three of which found at least partial evidence of lasting effects. A combination of a standard behavioural weight loss protocol and feedback from digital technology led to lasting weight loss after half a year of use (Pellegrini et al., 2012); a range of lifestyle-oriented interventions based on feedback had effects that were discernable even after two years after the single point intervention (Schulz et al., 2014); and delivering feedback to reduce eating rate led to a lasting decrease in weight after a
year of use, which was still discernable six months after intervention completion (Ford et al., 2009).

Contrarily, in a study of thirteen households that involved an in-home display of energy use, Quintal, Pereira and Nunes (2012) found no significant effects of display use on energy consumption even after a full year. However, this lack of findings may be due to a lack of control conditions and/or low statistical power, since descriptive data do point in the direction of a positive effect.

For behaviour change to take effect, however, sustained use of the intervention is needed: intervention adherence is known to be significantly correlated with intervention success (Burke et al., 2008). Only three studies looked into sustained use of the feedback technology. First, in a qualitative study on the use of health mash-ups translating information from different feedback sources into natural language, almost all participants used the intervention for the full 90 days of the project (Bentley et al., 2013). Contrarily, in a weight loss intervention (Pellegrini et al., 2012), 20% of participants stopped within 6 months; and Pereira, Quintal, Nunes, and Bergés (2012) found that even though they could report initial success, after four weeks interest in their feedback intervention on energy use was waning, with detrimental results on feedback effect. These latter two findings are in line with literature on sustained use of behaviour change interventions, which show a sharp decline in self-monitoring willingness after 10–14 days (e.g. Burke et al., 2008; Patrick et al., 2009) and a linear decline of the use of wearable technology which results in about 40% dropout within 12 months (Ledger & McCaffrey, 2014).

The effect of interpersonal and intrapersonal differences
Previous research has shown that not everybody benefits equally from feedback interventions. Both stable (traits) and dynamic (states) moderators are seen to influence feedback efficacy. Surprisingly, only one study in this review looked directly at the effect of demographic variables on feedback effectiveness. In an analysis of feedback on energy use in 2000 households, Vassileva et al. (2012) found that socio-economic factors such as income, age and type of housing interacted with the preferred medium of feedback delivery. Unfortunately, their work did not include the effect of socio-economic status on feedback effect.

In a similar vein, only a few studies took individual differences of any kind into account, be it motivation, self-regulatory capabilities, or personality traits. Bech-Larsen & Grønhøj (2013) found that people who consumed hardly any fruit benefited more from feedback than people who already consumed close to the desired target, suggesting a ceiling effect to feedback effectiveness that would cause underperformers to benefit more from feedback interventions than high performers. Similarly, Tasic et al. (2012) found that people who used a lot of water
Using feedback through digital technology to disrupt and change habitual behaviour

### Table 2.2
Effect sizes (reported or calculated) of interventions in the included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Effect Size</th>
<th>Dependent variable</th>
<th>n</th>
<th>Analysis Issues</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hurling et al., 2007</td>
<td>3.022</td>
<td>physical exercise</td>
<td>70</td>
<td>Other</td>
<td>b</td>
</tr>
<tr>
<td>Hoggan &amp; Brewster, 2010</td>
<td>2.5201</td>
<td>IQ training</td>
<td>9</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Stamopoulos et al., 2014</td>
<td>2.3528</td>
<td>buying domestic products</td>
<td>32</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Chang et al., 2008</td>
<td>2.129</td>
<td>Brushing teeth in children</td>
<td>13</td>
<td>c</td>
<td>a, b</td>
</tr>
<tr>
<td>Spelmezan, 2012</td>
<td>1.9604</td>
<td>snowboarding skill</td>
<td>10</td>
<td>a</td>
<td>a, b, c, a</td>
</tr>
<tr>
<td>Bruns Alonso et al., 2014</td>
<td>1.6101</td>
<td>toothbrushing stroke length</td>
<td>21</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Van Dantzig et al., 2013</td>
<td>1.188</td>
<td>sedentary behaviour</td>
<td>86</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Lee &amp; Dey, 2014</td>
<td>1.05</td>
<td>medicine adherence</td>
<td>12</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Brumby et al., 2011</td>
<td>1.0, 2.77</td>
<td>info-processing in car sim</td>
<td>24</td>
<td>a</td>
<td>a, c</td>
</tr>
<tr>
<td>Oshima et al., 2011</td>
<td>0.953</td>
<td>weight loss</td>
<td>56</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Wang &amp; Chenn, 2013</td>
<td>0.928</td>
<td>body massages, stretching in</td>
<td>39</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>computer use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>self-understanding in health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>behaviour, wellbeing</td>
<td>60</td>
<td>f</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>Bentely et al., 2013</td>
<td>0.887</td>
<td>driving eco-friendly</td>
<td>50</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Qian et al., 2011</td>
<td>0.835</td>
<td>walking pace</td>
<td>20</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Maltz &amp; Shinar, 2007</td>
<td>0.556, 0.317</td>
<td>keeping distance to next car</td>
<td>135</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Pellegrini et al., 2012</td>
<td>0.5198</td>
<td>weight loss</td>
<td>51</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Liu &amp; Pfaff, 2014</td>
<td>0.471</td>
<td>time not working, stress</td>
<td>30</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>Spring, 2012</td>
<td>0.43</td>
<td>weight loss</td>
<td>70</td>
<td>h</td>
<td>d</td>
</tr>
<tr>
<td>Donnez et al., 2008</td>
<td>0.4268</td>
<td>braking, accelerating, glancing</td>
<td>48</td>
<td>a</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in driving simulator</td>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>Bech-Larsen &amp; Grenhøj, 2013</td>
<td>0.381</td>
<td>fruit and veg. consumption</td>
<td>256</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Willis et al., 2010</td>
<td>0.322, 0.451</td>
<td>water usage, shower length</td>
<td>49</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Ford et al., 2009</td>
<td>0.293</td>
<td>eating rate in obese children</td>
<td>106</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Schulze et al., 2014</td>
<td>0.28, 0.18</td>
<td>health behaviours</td>
<td>5055</td>
<td>a</td>
<td>d</td>
</tr>
<tr>
<td>Alahmad et al., 2012</td>
<td>0.143</td>
<td>Home energy use</td>
<td>151</td>
<td>b</td>
<td>d</td>
</tr>
<tr>
<td>Kim et al., 2008</td>
<td>0.107</td>
<td>knowledge of peers’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sleeping behaviour</td>
<td>6</td>
<td>b</td>
<td>a, b</td>
</tr>
<tr>
<td>Quintal et al., 2012</td>
<td>0.052</td>
<td>electricity consumption</td>
<td>13*</td>
<td>e</td>
<td>a, b</td>
</tr>
</tbody>
</table>

1 – Effect size: Cohen’s d
2 – Analysis method: a = analysis of variance, b = t-test, c = nonparametric tests (e.g. Wilcoxon signed ranks), d = (Pearson’s) chi squared test, e = correlations and regression, f = descriptives only, h = other
3 – Issues: a = underpowered, b = no control condition, c = lacking conditions, d = other (such as self-report measures, self-selection, sample distribution issues)
4 – Field: a = design research, hci, engineering; b = health and psychology
* – number of households included in study, ** – number of classes included in study, *** – 120 in experimental condition, 15 in control condition
a – task priority x performance, no choice, b – task priority x performance, choice, c – distance, d – modality, e –
lenght, f – volume, g – tt, sequential, h – t2, simultaneous
for showering decreased their water use a lot more than people who used less. Wallenborn et al. (2011) found that men were more interested in the use of smart meters than women and indeed used them more.

Finally, the null result in research reported by Alahmad et al. (2012) might be seen as a further indication of ceiling effects in feedback interventions. If self-selection has a detrimental effect on the effectiveness of a feedback intervention, it might be that this is because participants are already performing the desired behaviour to the maximum possible extent.

The effect of feedback technology and properties
Feedback content factors (such as feedback sign, level of tailoring, and comparison level), the technology through which the feedback is delivered, feedback characteristics (such as timing, modality, frequency and duration), and the presentation of the feedback, all may influence the efficacy of feedback interventions. In this section, we first present results regarding feedback content, followed by results regarding feedback technology, characteristics and design.

For each study, we analysed the sign of the feedback, i.e. whether the digital technology delivered positive feedback (“You have exceeded your goal by 1,000 steps”), negative feedback (“you are still 1,000 steps short of your goal”) or neutral feedback (“you have managed 9,000 steps today”). Furthermore, we analysed the comparisons the digital technology made in delivering the data, i.e. comparing to past performance, peer behaviour, or abstract norms. Level of tailoring was not taken into account, because every study in the review included some form of tailoring.

Feedback sign. The vast majority of studies (55 out of 72) delivered feedback in such a way that both positive and negative feedback were possible, 4 studies only utilized feedback with a negative sign, and two studies only provided positive feedback. A further 12 studies provided neutral feedback, i.e. without any form of reference to performance goals or norms and therefore without sign. Two of these twelve studies combined neutral feedback for one dependent variable with signed feedback for another dependent variable. In one study, the feedback was described without detail, so no feedback sign could be established.

Only two studies directly compared positive and negative feedback. Both studies, which compared the effect of rewards and penalties on engagement (Jain et al., 2012), and the effect of positive with negative feedback on work pace interruptions (Liu & Pfaff, 2014), found a greater effect for positive feedback than negative. Moreover, the latter study found that negative feedback does indeed increase performance, but at the cost of a greater stress level.

Feedback Comparison. Different forms of comparisons can be made with feedback data. Current performance can be compared to past performance (historic
comparison), a social comparison with peers or unknown counterparts can be delivered, or performance can be compared to a norm or a goal (normative comparison). In this review, 52 studies made a normative comparison in their feedback. 18 studies gave historic comparisons (8 of which combining this with normative feedback, 1 with social feedback, and 2 with normative and social feedback), 7 studies used social comparison (3 of which in combination with other forms of comparisons). 7 studies delivered the data 'as is', without comparison. One study described the feedback without detail, so no information about comparison could be extracted.

Two studies contrasted different kinds of comparisons directly. Jain, Taylor, and Pescheira (2012) looked at the effect of normative and historic feedback comparisons in smart energy meters, finding that historic comparisons resulted in greater effect, whereas normative comparisons did not change energy use. In contrast, Sundramoorthy et al. (2011) found that normative, social and historic comparisons resulted in greater energy saving.

All in all, on the basis of the data extracted in this review, it is not possible to ascribe a more positive effect on feedback efficacy to a single strategy of comparison. This reflects findings in literature on feedback in general.

Feedback technology. To deliver the feedback, 16 studies utilized a mobile phone app, 11 studies used an in-home display – mostly for energy use monitoring –, in 9 studies, feedback was delivered using a website, and 7 studies used a computer or tablet application. Four studies provided participants with a wearable device capable of delivering vibrotactile feedback and three studies used a driving simulator. SMS text messaging, Facebook apps, and interactive public displays were used once. One study provided feedback both through a mobile phone app and a website. The largest category is that of the 'smart' devices, used in 18 of the studies. These devices often resemble generic household instruments, such as cutlery or scales, augmented with sensors and actuators. All but three studies derived the data for the feedback directly from the target behaviour; three studies relied on self-report for the generation of feedback.

Each feedback technology has particular characteristics that impact the overall experience of the user. The wearable vibrotactile devices could only deliver feedback in their own modality, concurrent with behaviour, and without possibilities for comparison to earlier results or performance of others. SMS text messages could only be delivered retrospectively, as they rely on technology with a time lag. However, technology choice was not associated with differences in effects on habit disruption or change; positive results as well as null findings were spread evenly across technologies. Unfortunately, none of the studies in the analysis directly compared different technological channels.
Chapter 2

Feedback timing. Of the reviewed studies, 20 delivered retrospective feedback, i.e. feedback after the behaviour had been performed. 52 studies delivered concurrent feedback, i.e. during behaviour performance. Two studies offered both forms for different behaviours, without a direct comparison. One study (Donmez et al., 2008) directly compared the effectiveness of feedback timing on behaviour. In this study, a combination of retrospective and concurrent feedback yields greater effect than separate timing strategies, because of the additional informational benefit offered by recurrent feedback on top of the direct intervention in behaviour offered by concurrent feedback. Furthermore, Tulusan, Staake, and Fleisch (2012) find that users of their eco-driving support application prefer direct, concurrent feedback over retrospective feedback: the efficacy of the application is significantly predicted by the usage of the direct feedback delivered by the app, but not by retrospective, indirect feedback.

Feedback modality. Of the studies included in the review, 58 studies offered visual feedback only, one offered auditory feedback only, and 8 studies used tactile feedback only. Five studies directly compared the effectiveness of different feedback modalities, two of which contrasted visual with auditory feedback, one study contrasted auditory with tactile feedback; one study contrasted visual with tactile feedback, and one study compared three feedback modes: visual, auditory and tactile. Studies comparing tactile feedback with other modalities found this modality more effective when aimed at changing motor skills (Epstein et al., 2012; Maltz & Shinar, 2007) and when disruptiveness mattered. Generally, tactile feedback was found to be less disruptive in other tasks compared to auditory feedback, which in turn is less disruptive than visual feedback. A reverse pattern can be observed in the amount of detail that can be communicated through different feedback modalities: visual feedback can be more detailed than auditory, which can offer more detail than tactile feedback (Hoggan & Brewster, 2010). One study (Epstein et al., 2012) reported an effect of feedback modality on the durability of the achieved behaviour change: sitting posture was changed beneficially through visual feedback, but only the addition of tactile feedback on optimal posture led to lasting effects.

These studies serve as an indication that the optimal selection of feedback modality not only depends on the targeted behaviour, but also on the amount of disruption that a given task allows and the necessary detail of the feedback. More evidence to support this assumption is needed.

Three studies support the assumption that multimodal feedback is more effective than single-mode feedback (Hoggan & Brewster, 2010; Lieberman & Breazeal, 2007; Qian et al., 2011). In these cases, the increased effect mostly lies in additional strengths of different feedback mode, for example tactile feedback in smartphones being more effective in noisy areas and auditory feedback more
Using feedback through digital technology to disrupt and change habitual behaviour

effective in silent areas. Maltz and Shinar (2007) tested the concurrent application of visual and auditory feedback in driving behaviour and found no beneficial effect of multimodal feedback, leading to the conclusion that auditory feedback is most effective for driving behaviours and other modalities do not add further improvement.

**Feedback frequency and duration.** The greater part of included studies (67 out of 72) used either continuous or on-demand delivery of feedback, which means almost all studies made use of the possibilities digital technologies offer in delivering the feedback as soon as possible. No studies compared the effect of different delivery frequencies directly. From the current literature, no conclusions can be drawn on the effectiveness of feedback frequency on feedback impact.

The duration of the feedback intervention differed from a single trial to one year. Those studies reporting lasting intervention effects had durations of six months (Pellegrini et al., 2012; Schulz et al., 2014), and one year (Ford et al., 2009). However, there is an obvious confound of intervention length with the type of behaviour targeted, because not every habitual behaviour is equally difficult to change, with periods needed for change ranging from a few weeks to behavioural vigilance without time limit (Lally & Gardner, 2013). Therefore, a single standard of ideal feedback intervention duration and frequency seems conceptually impossible.

**Feedback presentation: usability and aesthetics.** Three studies considered the effect of visual design on feedback effectiveness directly. All three found some explorative indication that design and aesthetics matter for feedback acceptance, use of the feedback device and feedback impact. One study (Consolvo, McDonald, & Landay, 2009) provides a very useful list of directives for the design of feedback presentation. The authors state that feedback should be abstract and reflective, unobtrusive and public, aesthetically pleasing, positive, controllable, trending/historical in comparison, and comprehensive. Two studies (Nakajima & Lehdonvirta, 2013; Rodgers & Bartram, 2011) described how heightened abstraction and aesthetic pleasantness seem to come at a cost in terms of usability and comprehension.

**Other insights**

Close scrutiny of all reviewed studies revealed a couple of noteworthy additional themes that were not detected in the analysis of existing literature that led to the hypotheses posed in this review.

One additional theme that emerged is the role of disruption in feedback efficacy. Feedback can play a role in habit change by disrupting the automatic response to a cue. However, this disruption may also cause a task to be abandoned or otherwise disturb task resumption (Bittner & Zondervan, 2015). The amount of disruption therefore needs to be carefully tailored to break the automatic cue-response-chain without abandoning the task altogether. In this analysis, two studies mentioned
the role of disruptiveness on feedback effect. As mentioned above in the section on feedback modality, a study of feedback delivered by a mobile game with different feedback modalities (Hoggan et al., 2009) exhibited an interaction between feedback modality, disruption, and richness of the feedback. Interestingly, one study (Liu & Pfaff, 2014) showed how feedback can also be used to facilitate the resumption of tasks after disruptions.

Another important insight is that the amount of integration of feedback in other areas of behaviour, such as usage of similar interventions or sharing behaviour on online social networks, might be a strong predictor of feedback effect. Wallenborn, Orsini and Vanhaverbeeke (2011) found that when energy monitors are not integrated in pre-existing practices, the information quickly disappears into background noise like with any other new appliance. A study by Jain, Taylor & Pescheira (2012) had a similar finding in a study of the usage of an interface providing feedback on energy consumption. Bentley et al. (2013) found similar patterns in the effect of health mashups. When participants used an app that integrated Fitbit activity tracking data with weight, food intake, sleep etcetera, sustained use of the feedback technology increased.

This notion of integration is an interesting concept that needs further exploration. Indeed, relevant theories that explain the effectiveness of feedback on behaviour change, such as Social Cognitive Theory (Bandura, 1994) or Control Theory (Carver & Scheier, 1985; Kuhl, 1985), suggest that behaviour change is most likely if feedback is not delivered on its own, but embedded in larger interventions with clear target behaviours and action plans. This notion is also backed up by considerable evidence from original research (e.g. Avery et al., 2012; Godino et al., 2013; Sniehotta, Nagy, Scholz, & Schwarzer, 2006) and reviews (Dombrowski et al., 2012; Gardner et al., 2010).

Wallenborn, Orsini and Vanhaverbeeke (2011) noted that wasteful behaviour in energy use can arise from role perception (“a good parent always gets the laundry clean and therefore washes at 90˚C”) and different levels of technical insight in families might lead to conflicts about the performance on feedback. This gives insight in how social interactions influence feedback effect. Feedback on performance spurs discussion with family members and others, which may in itself lead to behaviour change or even conflicts and role clashes. Similar effects are reported by Kappel and Grechenig (2009) when they mention positive social effects of their device that reports water usage in the shower: “A couple used to argue that one of them always took longer in the shower and (...) used more water. (...) (T)hey learned that the woman used only half as much water, even though she spent more time in the shower. This discovery stimulated the man to further reduce his own water consumption. In another household the child (11 yrs.) triggered discussions about the water consumption, because he used much less water than his parents. This
Using feedback through digital technology to disrupt and change habitual behaviour

stimulated his mother to begin reducing her own consumption (…).” Nakajima & Lehdonvirta (2013) and Katzeff et al. (2013) found similar results in an intervention aimed at (respectively) children’s tooth brushing and energy use in the office.

**Conclusion**

This review shows that in the 72 studies we analysed, feedback delivered through digital technology is generally effective in disrupting habitual behaviour. However, the current literature does not provide enough evidence to support the hypothesis that feedback through digital technology leads to lasting behaviour change. Furthermore, little is known about factors that facilitate sustained use of digital technology, intra-personal and inter-personal moderators of feedback efficacy, and the effect of feedback characteristics.

This review makes clear that feedback through digital technology has the potential to disrupt undesired habits. Therefore, such feedback can be seen as a potentially reinforcing ingredient for any intervention aimed at habit change. This work offers support for Quantified Self-solutions, which may indeed lead to healthier, more eco-friendly behaviours; it also supports the notion that delivering feedback through digital technology may heighten the chances of conscious scrutiny for a broad range of deeply engrained, undesirable habits. Our analysis shows this finding is consistent across feedback technologies: feedback delivered through a broad range of technological channels appears to succeed in disrupting undesired habits.

However, the possibilities of using feedback through digital technology for sustainable habit change have yet to be proven. Particularly, the durability of the feedback effect on habitual behaviour is as yet unclear. Those few studies that included longitudinal measurements generally found sustainable effects of feedback on behaviour, but the greater part of the studies only measured effects right after the intervention. To prove the hypothesis that feedback through digital technology actually enables users to change their behaviour, more evidence on whether the use of the digital technology leads to lasting effects is necessary.

To ensure the occurrence of behaviour change, intervention designers must make sure their technology is accepted by its users and used long enough to warrant habit change. Existing literature (e.g. Ledger & McCaffrey, 2014) suggests that technological feedback solutions are often to be discarded after initial use. Unfortunately, methods to maintain engagement with a technology over time remain unclear.

The role of moderating traits and demographic factors also remains understudied. Very little is known of the interplay of traits and states on the one hand,
and feedback properties such as feedback sign, comparison, and delivery mode on the other. Similarly, the effect of different feedback properties such as timing, modality, frequency and duration, have not yet received the attention needed to draw any conclusions on their impact on feedback effect. This suggests that we cannot yet tell whether changes in behaviour can really be attributed to the digital technology and its feedback, or that these are merely functioning as some sort of lens through which only well-motivated and capable individuals manage to focus their behaviour-changing endeavours.

Although this review provides evidence for the effect of feedback through digital technology on disrupting habitual behaviour, this review also demonstrates that research into such effects has only just started. Because of the explorative, descriptive nature of a great part of the included studies, there are limits to the conclusions that can be drawn from this review. The majority of the included quantitative studies, 33 out of 50, did not report any control measures for demand characteristics or other experimental artefacts, e.g. through well-balanced experimental designs. Furthermore, 22 out of 30 quantitative studies with statistical analysis were statistically underpowered, which seriously reduces the validity of any conclusions drawn from those studies. As a consequence, only a part of the 72 original studies in this review (14 quantitative studies and 21 qualitative studies) were described in a way that proves enough methodological rigor to act as a source for direct evidence. The literature would benefit greatly from well-performed additional research on the effect of feedback through digital technology on habitual behaviour, be it field studies or lab work, with good active controls for experimental artefacts and ample statistical power.

Moreover, it remains unknown how many studies did not make the literature because the desired effect could not be shown or no support was found for the original hypothesis. The great majority of studies in this review found a positive effect of feedback on habit disruption, much more so than in similar analyses (e.g. Kluger & DeNisi, 1996, who find a 66% success rate). The field (and science in general) would greatly benefit from measures aimed at reducing publication bias, such as pre-registering studies, to provide insight into how many ‘failed’ studies end up in the proverbial file drawer (Franco, Malhotra, & Simonovits, 2014).

The review also shows the merit of combining quantitative research with good qualitative and explorative research. It is paramount that theories of behaviour change are supported by well-designed trials, but important insights such as the influence of social interaction on the effects of feedback delivered by digital technology would not easily show up in even the most well-set up quantitative research.
Using feedback through digital technology to disrupt and change habitual behaviour

Further research
All of these areas provide ample possibilities for further research. The broad range of dependent variables and feedback technologies limit the validity and generalizability of the findings in this review. However, the results presented here may serve as a basis for further studies and analyses.

One such analysis could examine which behaviours are most likely to benefit from feedback delivered through digital technology. Intuitively, the hypothesis that feedback does not affect every habitual behaviour equally seems plausible, but evidence is lacking. Similar questions arise when the different technologies are taken into view. Different technologies offer different possibilities for feedback modality and other properties. It seems plausible to assume that these differences influence efficacy, but this does not follow from the results of this review. Particular attention should be paid to the level of disruption of the feedback.

Evidence (Bittner & Zondervan, 2015) suggests that feedback may disrupt tasks in such a way that this leads to task abandonment. Some feedback modalities (visual) are clearly more disruptive than others (vibrotactile, auditive). The effects of feedback disruptiveness on sustained performance warrant further scrutiny.

Factors moderating the sustained use of technological solutions are another area that deserves our attention. Without use, we cannot expect technology to have any effect on behaviour. User experience, usability, and design can be thought of as moderating factors on the effect of feedback, but as yet this hypothesis lacks support. Intuitively, and from what little evidence that exists (e.g. Ludden, Van Rompay, Kelders, & Van Gemert-Pijnen, 2015), one would reason that clunky designs are unlikely to get used, with detrimental consequences. Therefore, we see the lack of focus on usability in this research field as a serious problem. Similar focus is needed on other factors influencing the lasting use of technological feedback solutions. Is a high motivation essential? Do certain personality characteristics facilitate sustained use, and what is the effect of feedback characteristics? All these questions need an answer.

Another example of an area of interest that deserves further scrutiny is the effect of personality traits and states such as initial motivation and self-efficacy on feedback impact. Literature suggests that high initial motivation and self-efficacy increase the impact of feedback on habitual behaviour. However, results from studies in this study suggest a ceiling effect. A well-set up experimental design could shed light on the effect of initial motivation and perceived self-efficacy on the effect of feedback on habits.

A similar question remains about the effect of feedback sign. In this review, the greater part of the studies provided feedback in such a way that both positive and negative feedback was possible. Unfortunately, this makes it impossible to test an interesting hypothesis, i.e. concerning the interaction between feedback sign and
regulatory focus – the tendency to approach positive impulses and avoid negative ones. Van Dijk and Kluger (2004; 2011) suggest that in a prevention focus (avoiding negative consequences), negative feedback should have more effect, whilst in a promotion focus (approaching positive consequences), positive feedback should have more effect. Hattie and Timperley (2007) however, find in a meta-analysis that positive feedback should always lead to more effect than negative feedback. This issue is particularly relevant to feedback delivered through digital technology, which by nature is capable of delivering both signs, depending on individual performance. Is feedback more effective in a prevention focus as long as goals are being reached, and does it lose its effect when goals are too hard - and similarly, is feedback more effective in a prevention focus as long as goals are not reached yet? Further research could give valuable insights in when feedback through digital technology has the most effect.

In a similar vein, the optimal choice of feedback properties in such a way that feedback is delivered concurrently with behaviour in a continuous or on-demand manner, and data gathering for the feedback takes place automatically without the need for self-report measures, should intuitively lead to an enhanced feedback efficacy. This hypothesis, however, remains unsubstantiated. Subjects of similar interest that have not been researched in a controlled manner at all are the active integration of feedback through digital technology within more complex interventions, and the social effects of digital technology. In real-life situations, feedback is not delivered in a vacuum, but plays a role in a social practice. Users will interact with friends, family and others about the received feedback, the attainability of goals, and the use of the artefact that delivers the feedback. The effects of feedback integration and social practices on feedback efficacy are in urgent need of research.

Further research into the effectiveness of feedback interventions to disrupt habits, personal differences in feedback efficacy, and the effect of applying different feedback characteristics, might not only enhance our knowledge on how habits might be changed. Such research would also serve as a basis for intervention developers and designers to inform the design of more effective behaviour change products. The ubiquity of Quantified Self-solutions and health-related apps on smartphones show a great level of acceptance of this kind of intervention. The public is generally ready and willing to embrace such interventions. Badly set-up products without a base in scientific evidence might do lasting damage to the benevolent reception feedback interventions currently receive. But well-designed, evidence-based solutions can be expected to have a great impact on our well-being and on the proliferation of sustainable behaviour. Feedback through digital technology as an intervention strategy to change undesirable habitual behaviour offers great chances for healthier and more sustainable living that should not be wasted.
Using feedback through digital technology to disrupt and change habitual behaviour
Chapter 3: Determinants for Sustained Use of an Activity Tracker: An Observational Study

Abstract

Feedback from wearables such as activity trackers has the potential to encourage daily physical activity. To date, little research is available on the natural development of adherence to activity trackers or on potential factors that predict which users manage to keep using their activity tracker during the first year (and thereby increasing the chance of healthy behaviour change) and which users discontinue using their trackers after a short time. The aim of this study was to identify the determinants for sustained use in the first year after purchase. Specifically, we look at the relative importance of demographic and socioeconomic, psychological, health-related, goal-related, technological, user experience–related, and social predictors of feedback device use.

A total of 711 participants from four urban areas in France filled out three web-based questionnaires: at start, after 98 days, and after 232 days to measure the aforementioned determinants. Furthermore, for each participant, we collected activity data tracked by a Fitbit tracker for 320 days. We determined the relative importance of all included predictors by using Random Forest, a machine learning analysis technique. The data showed a slow exponential decay in Fitbit use, with 73.9% (526/711) of participants still tracking after 100 days and 16.0% (114/711) of participants tracking after 320 days. On average, participants used the tracker for 129 days. Most important reasons to quit tracking were technical issues such as empty batteries and broken trackers or lost trackers (21.5% of all Q3 respondents, 130/601). Random Forest analysis of predictors revealed that the most influential determinants were age, user experience–related factors, mobile phone type, household type, perceived effect of the Fitbit tracker, and goal-related factors. We explore the role of those predictors that show meaningful differences in the number of days the tracker was worn.
Introduction

The Effect of Activity Tracker Usage on Physical Activity

One of the biggest threats to our health is physical inactivity, which is considered to cause 6% of deaths globally (WHO, 2017). Too little physical activity plays a role in a range of debilitating conditions such as cardiovascular diseases, diabetes mellitus type II, chronic obstructive pulmonary disease, and some forms of cancer (Pedersen & Saltin, 2006; Warburton, Nicol, & Bredin, 2006). The American Heart Association endorses 10,000 steps a day or 30 min of moderate-intensity physical activity (e.g., brisk walking) for at least 5 days a week as guidelines to improve health and reduce health risk (Tudor-Locke et al., 2011; Warburton et al., 2006). Unfortunately, many people fail to meet these criteria (Lee et al., 2012).

Behaviour change toward more physical activity might greatly benefit our health. Unfortunately, for many people, their physical activity is a deeply engrained habit (Kremers & Brug, 2008; Phillips & Gardner, 2016). Choosing physical activity over inactivity tends to occur outside awareness (Phillips & Gardner, 2016). This lack of conscious scrutiny is one of the main reasons sedentary habits are difficult to change; we are not always adept in monitoring our own behaviour, especially not when this behaviour is executed unintentionally (Wilson, 2002). It is not surprising, therefore, that people tend to overestimate their physical activity (Godino et al., 2014; Vooijs et al., 2014). Supporting our self-monitoring abilities by providing us with timely and relevant feedback on our behaviour has proven a successful strategy to disrupt the automaticity of deeply engrained habitual behaviours such as inactivity and make them available for conscious scrutiny (Hattie & Timperley, 2007; Hermsen, Frost, Renes, & Kerkhof, 2016; Verplanken & Wood, 2006).

In recent years, numerous interactive and mobile technology solutions to encourage physical activity have arrived in the form of devices that are able to directly monitor our physical activity through a range of sensors. The information thus gathered can be applied by automatically providing the user of the device with behaviour change techniques (BCTs) from the monitoring cluster (Michie et al., 2013): timely feedback on their own behaviour and the possibility to self-monitor behaviour and its outcomes. Furthermore, dashboard applications often encourage (but hardly ever enforce) a range of secondary BCTs: goal setting, the review of behavioural goals and their outcomes, and social comparison and support.

Such activity trackers are an increasingly popular way to promote physical activity. In 2012, a survey showed that 69% of adults in the United States tracked their physical activity.
at least one health behaviour using some sort of tracking device, and 14% of US citizens owned a specialized activity tracker of some sort (Fox & Duggan, 2013). Of those who did track a health behaviour, roughly half indicated that tracking changed their overall approach to maintaining their health (ibidem).

The effect of using activity tracker technology on physical activity is well established for a range of populations (Bravata et al., 2007; Jennings et al., 2016; Thorup Msn et al., 2016; Wang et al., 2015); however, a crucial ingredient for lasting effects of behaviour change interventions in general is the sustained use of the intervention (Hermsen, Frost, Renes, et al., 2016). Unfortunately, even though there is a growing body of research utilizing activity trackers, there is as yet little research available on sustained use of such devices. Anecdotal evidence, as well as what little evidence that is available (Wang et al., 2015), suggests activity trackers may have a poor record when it comes to sustained use, as they are easy to switch off, ignore, lose, or neglect. Furthermore, there is to date no research available that sheds light on which users manage to stick to using their activity tracker during the first year (and thereby increasing the chance of healthy behaviour change) and which users stop using their trackers after a relatively short time. This chapter attempts to add to our knowledge of the sustained use of activity trackers and factors that predict this sustained use.

**Potential Determinants of Tracker Use**

On the basis of evidence from prior research on the effect of feedback interventions on habitual behaviours (e.g. Fischer, 2008; Hermsen, Frost, Renes, et al., 2016), there is a broad range of factors that might influence sustained use and efficacy of activity trackers.

First, tracker technology may play a crucial role in sustained use. Trackers may be abandoned because of empty batteries, with the perceived cost of replacement too high or too cumbersome (Harrison, Berthouze, Marshall, & Bird, 2014). Apart from technical failures, actual or perceived characteristics of the tracker may fit user expectations. The user experience and ease of use (Clawson, Pater, Miller, Mynatt, & Mamykina, 2015; Lazar, Koehler, Tanenbaum, & Nguyen, 2015), functionality or lack thereof (Clawson et al., 2015), the possibility to upgrade toward a newer device (ibidem), aesthetics and form (Harrison, Marshall, Bianchi-Berthouze, & Bird, 2015), perceived accuracy (ibidem), and perceived fit between device and self-image (Lazar et al., 2015) are all reasons to either abandon the tracker or to keep using it. Furthermore, the data delivered by the tracker must fit participants’ needs (ibidem). Finally, computer literacy, or the perceived self-efficacy in using digital devices, is known to affect sustained use (e.g. Couper et al., 2010; Peels et al., 2012: higher more than lower).
Socioeconomic status markers such as education (e.g. Couper et al., 2010; Geraghty, Torres, Leykin, Pérez-Stable, & Muñoz, 2013: higher more than lower) and employment (e.g. Al-Asadi, Klein, & Meyer, 2014; Habibović et al., 2014: higher more than lower), age (e.g. Bossen, Buskermolen, Veenhof, De Bakker, & Dekker, 2013; Couper et al., 2010; Davies et al., 2012: older more than younger) and gender (e.g. Couper et al., 2010; Geraghty et al., 2013; Wanner, Martin-Diener, Bauer, Braun-Fahrlander, & Martin, 2010: women more than men) are known to influence sustained use, as are psychological traits such as inhibitory strength and the capacity for self-regulation (Baumeister & Heatherton, 1996; Braverman, 2008; Kuhl, 1985).

Personal health-related factors may very well influence the sustained use of the activity tracker; poor health decreases perceived self-efficacy (Grembowski et al., 1993; Strecher, Devellis, Becker, & Rosenstock, 1986), which is known to influence sustained use (Cugelman, Thelwall, & Dawes, 2011). Low mood, stress, sleep disturbances, and other markers of mental health, are also known to decrease sustained use (Bossen et al., 2013; Christensen, Griffiths, & Farrer, 2009).

Goal-setting is generally seen as a promising strategy to increase the use of physical activity interventions (Wang et al., 2012; Wilson & Brookfield, 2009). Strong, clear goals and motivation to fulfil these goals (Bossen et al., 2013; Couper et al., 2010; Davies et al., 2012) increase the chance of sustained tracker use. Achieving these goals, or at least displaying a performance level that could lead to achieving previously-set goals, can provide a further boost to initial motivation and perceived self-efficacy, increasing the chances of sustained tracker use. However, the fulfilment of a set goal may also lead to device abandonment, because users feel they no longer need the tracker (Murnane, Huffaker, & Kossinets, 2015).

Furthermore, behaviour change theories (e.g., social cognitive theory (Bandura, 1977) and control theory (Kuhl, 1985)) suggest that behaviour change is most likely if feedback is not delivered on its own but embedded in larger interventions with clear target behaviours and action plans. Combined use of the activity tracker with other health apps, participation in a therapeutic regime, and use of the app and Web-based platform that accompany the activity tracker may be seen as an operationalization of this concept of integration. Overall, we expect users with strong goals and high integration of their tracking behaviour in other health-related practices to have a higher chance of sustained tracker use, especially when these users manage to achieve their performance goals.

Feedback properties such as timing, duration, frequency and sensory modality (cf. Fischer, 2008), and user experience (e.g. Tractinsky et al., 2000) are known to influence the efficacy of the feedback intervention, both directly and through perceived usability and agreeableness. Similarly, feedback properties (Michie & West, 2016) and user experience-related factors are known to affect the uptake and sustained use of physical activity trackers (Canhoto & Arp, 2017). We expect
users with greater liking of the tracker and its accompanying online tools to have a higher chance of sustained tracker use.

Activity tracking is often social and collaborative instead of individual and personal (Fritz, Huang, Murphy, & Zimmermann, 2014; Harrison et al., 2015; Rooksby, Rost, Morrison, & Chalmers, 2014). Social interaction is known to improve adherence to physical activity interventions in general (DiMatteo, 2004). We therefore expect users that share their tracking data with peers or relatives to have a higher chance of sustained tracker use.

Sample Size and Duration in Previous Research on Activity Trackers

Current research into determinants of activity tracker use typically makes use of small test populations, ranging from 7 to 31 participants (Fritz et al., 2014; Harrison et al., 2015; Lazar et al., 2015; Munson & Consolvo, 2012; Rapp & Cena, 2015; Rooksby et al., 2014; Shih, Han, Poole, Rosson, & Carroll, 2015), which limits the possibilities to reliably investigate quantitative measures of determinants of device use. When larger samples have been tested (e.g. Clawson et al., 2015, n = 1561, and Wilson & Brookfield, 2009, n = 256), only a small number of determinants were included. Furthermore, adherence studies generally covered only a very short period, that is, 2 months or less (e.g. Lazar et al., 2015; Munson & Consolvo, 2012; Rapp & Cena, 2015; Shih et al., 2015). Only one study (Gouveia et al., 2015) tested sustained use over a period of up to 10 months. However, this study did not evaluate potential determinants for adherence.

This study attempts to contribute to bridging this knowledge gap by looking into factors predicting sustained use in the first year after purchase. Specifically, we look at demographic and socioeconomic, psychological, health-related, goal-related, technological, user experience–related, and social predictors of feedback device use and their predictive power in determining which participant is most likely to continue using the device.

Methods

Study design

This study was initiated by IDS Santé Inc (Paris, France), a full-service communication agency aimed at the health sector and specializing in prevention and health education and executed from June 2013 until winter 2014 as a project called MySantéMobile. A total number of 1000 participants were recruited in France via a (free) newspaper from four French cities (Bordeaux, Lille, Montpellier, and Lyon). Each participant received an activity tracker and was requested by email to fill in three Web-based questionnaires (June 2013, August 2013, and January 2014).
completion of the study, the full raw dataset was transferred for independent and retrospective analysis to the authors of this chapter.

To establish which set of the included predictors best explains the use and non-use of this activity tracker in the dataset, we adopted the Random Forest method, a machine learning approach (Breiman, 2001). This approach enables identification of predictors that explain large portions of variance while minimizing the risk of overfitting, which is likely to occur when performing a regression analysis with a large set of predictors (Strobl, Malley, & Tutz, 2009). Furthermore, this approach is also capable of detecting nonlinear relationships and higher-order interactions between predictors.

Activity Tracker
The activity monitor used in this study, the Fitbit Zip, is a small (2.9 cm x 3.6 cm x 1 cm) consumer device that tracks activity through counting steps. The Zip is worn as a clip-on device on the waist or elsewhere where it can be easily clipped onto clothing. On the device screen, the Zip displays the number of steps taken on the current day, and, after pressing a button on the device, displays the distance covered on the current day, active minutes, the time, an approximation of calorie expenditure, and feedback in the form of a happy, neutral, or unhappy smiley.

Research (Evenson, Goto, & Furberg, 2015; Kooiman et al., 2015) shows that the reliability and validity of the Fitbit Zip activity monitor is high, with little error in the number of registered steps, both in laboratory conditions and in daily life.

Participants

Recruitment
Participants were recruited through a newspaper article, published on the 14th of May 2013, in free newspapers in France. 1000 participants were selected using the following inclusion criteria: living in one of the four eligible cities (Montpellier, Lyon, Lille, and Bordeaux); at least 18 years of age; and owning a smartphone or computer compatible with Fitbit. Of those 1000, 929 received a Fitbit Zip activity tracker and took part in the study.

Data Acquisition
In the first week of June (2013), all eligible participants were invited to fill out a Web-based questionnaire by email. This questionnaire was presented through the LimeSurvey platform and covered sociodemographics, device usage, tablet/ phone brand, self-reported tracker use, use of other health apps and devices, health, exercise, and diet. All questionnaires used in this study are available as an online appendix (see list of supplementary materials, Appendix 2). Approximately two
weeks after filling in this questionnaire, the participant received their Fitbit Zip tracker by mail. Participants received their Fitbit Zip tracker free of charge.

Upon dispatch of the Fitbit trackers, participants received an email giving them instructions on how to install and use the Fitbit, how to synchronize data and how to authorize MySantéMobile in acquiring their data through the Fitbit API. Instructions were also provided on the MySantéMobile website. Participants then had to give permission to MySantéMobile to read their activity data through the Fitbit API. Participants who did not give permission received reminder phone calls and emails.

A second questionnaire was sent out by email on 23 August 2013 (after 98 days). The third questionnaire was also sent out by email, on 7 January 2014 (232 days). Participants who did not fill out the questionnaire received a reminder email after two weeks. Participants received no incentive other than a free activity tracker. At the end of the data acquisition period, all participants received an overview of the study results.

Table 3.1
Participant characteristics at Q1

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>330 female, 381 male</td>
</tr>
<tr>
<td>Marital status</td>
<td>Single: 240, Couple: 332, Single parent: 38, Family: 272, Other: 54</td>
</tr>
</tbody>
</table>

Participant Selection
Since the selected analysis method does not allow missing values, only data from those participants who completed their questionnaires could be used. Of the 929 participants originally approached to take part in the study, 711 participants (76.5%) completed the first questionnaire and gave permission to MySantéMobile to read their activity data through the Fitbit API. Data collection using the Fitbit took place from 20 June 2013 to 13 May 2014 (327 days). Of this group of 711 participants, a total number of 575 participants (80.8%) completed the second questionnaire (August 2013) and 542 participants (76.2%) completed both the second and the final questionnaire (January 2014).

Measures
The total number of days on which the device was worn was used as the primary outcome measure (adherence to using the wearable for self-tracking). We only had access to data that were synchronized with a personal computer or mobile app. However, the Fitbit Zip stores steps data for 30 days, therefore, we assume
most active users will synchronize their data within this time window. For ease of 
interpretation, we will speak of “using” or “wearing” the Fitbit. However, note that 
our measure may somewhat underestimate the number of days the Fitbit was worn. 
Furthermore, we calculated the average amount of steps taken by each participant 
on those days the tracker was used.

Questionnaires and Item Selection
Three questionnaires (Q1, June 2013; Q2, August 2013; Q3, January 2014) were 
sent out to the participants. A complete overview of all three questionnaires, with 
the exact questions (translated into English), and the response scales used for 
each question, is available as an online appendix (see Supplementary Materials, 
Appendix 2).

From these questionnaires, we selected for our analysis those items that (1) 
matched the potential determinants for sustained use of the tracker outlined in the 
introduction of this chapter, and (2) met with our requirements for item validity.

On the basis of our analysis of potential determinants for sustained use, we 
cluded the following items from the questionnaires in our analysis:

Demographical and socioeconomic factors: age, gender, place of residence, 
household size and household composition, profession, and education (all in 
questionnaire 1 (Q1).

Psychological factors: general mood (all questionnaires); specific scores on 
affective situation (sadness, gaiety), stress (calmness, stressfulness), energy 
(energy level, tiredness), and sleep quality (all in all questionnaires); big five 
opersonality traits (openness to experience, conscientiousness, extraversion, 
agreeableness, and neuroticism; plus, rebelliousness, health-mindedness, and 
independence (all in Q3).

Technological factors: synching platform type (smartphone, tablet, computer), 
operating system - iOS or Android (all in Q1), use of other health applications (Q1), 
experience with technology (Q3)

User experience: perceived utility, enjoyableness, intrusiveness, modernity, fun, 
reliability, simplicity, inconvenience, correspondence to needs, beauty, robustness, 
and cumbersomeness of the activity tracker (all in Q2); exactness, detail, clarity, 
credibility, confidence, insight, perceived efficacy (all in Q3)

Health-related factors: body mass index (all questionnaires), smoking (Q1), 
pregnancy (Q1), diet (Q1), medical treatment status (Q1), activity in sports (Q1), and 
sports together with others (Q1).

Predefined participant goals and perceived goal achievement: increasing 
activity, improving sleep, quitting smoking, diagnosing or improving diet, diag-
osing behaviours, losing weight, and improving stamina (all in Q1); for each goal, 
the perceived achievement of the goal was measured (Q2 and Q3).
Chapter 3

Social factors: whether participants talked about the tracker sharing use with family, friends, colleagues, teams and clubs; sharing data on the Internet through social media, blogs, Twitter, websites, forums, and mailing lists (Q2 and Q3).

Questionnaire Validity
Because of the history of this study, which started as groundwork for a publicity campaign for a communications agency, the questionnaires used in this study have not been constructed in such a way that meets the current standards for validity. To evaluate the validity of the three questionnaires used in this study and to determine which items were of high enough standard to include in our analysis, we compared each question with current, well-validated standard approaches in scientific literature. The complete result of this analysis is included in Appendix 2 (see Supplementary Materials). For each item, under “remarks,” the validity evaluation is listed. Generally, our evaluation showed that the greater part of the questionnaire items survives rigid scrutiny and satisfies scientific criteria. However, the validity of four items, one item on digital proficiency and three items on psychological traits (rebelliousness, independence, and health-mindedness) could not be satisfactorily assessed. Results for these items should be used with caution.

The greater part of the questionnaire consisted of single-item measures. Single-item measures can be eminently usable (sometimes even more so than multiple item measures) when the attribute (e.g. attitude, frequency) is concrete and singular (i.e. not consist of multiple facets) and when the object of the item (e.g. brand, product) is concrete (Bergkvist & Rossiter, 2007). For most of the items, this is the case; see Appendix 2 for an overview. Three exceptions occurred, which are as follows: items regarding emotional well-being, items regarding psychological traits, and items regarding user experience.

The first group of items that do not have a concrete object are about emotional well-being. However, single-item assessments of emotional well-being are often used in large-scale surveys (see Cooke, Melchert, & Connor, 2016 for an overview) and have been shown to perform quite well compared with multiple-item scales (e.g. Jovanović, 2016). We can therefore probably conclude that these single-item self-report measures are a sufficiently valid measure for the purpose of this chapter.

The second group of items that do not have a concrete object concern psychological traits. In Questionnaire 3, the big five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) are measured using the French translation of the TIPI questionnaire (Gosling, Rentfrow, & Swann, 2003). This questionnaire has been well validated.

A third group of items address the user experience of the Fitbit. User experience can be defined as a person's perceptions and responses that result from the use and/or anticipated use of a product, system, or service (International Organization
for Standardization, 2010). In recent years, approximately a hundred different measures have been developed (Vermeeren, Law, Roto, Obrist, Marianna Hoonhout, & Väänänen-Vainio-Mattila, 2010). User experience evaluations generally consist of questions addressing some (but never all) of the following concepts: timeliness, adaptability, comfort, opacity, efficiency, immersion, intuitiveness, ease of use, usefulness, interaction, controllability, clearness, completeness, identity, novelty, originality, fun, stimulation, valence, connectedness, attractiveness, beauty, and trust (Winter, Schrepp, & Thomaschewski, 2015). Currently, no questionnaires or other measures exist that address all of the aforementioned concepts. In this study, comfort, ease of use, novelty, fun, valence, attractiveness, trust, and invasiveness have been measured.

Concepts regarding user experience can be subdivided in three categories (Hassenzahl, 2003): pragmatic qualities (usability-oriented; first 13 concepts [timeliness-completeness]), hedonic qualities (identity-connectedness), and general concepts, mainly focused on attraction. Principal component analysis showed that the response patterns for the 22 user experiences-related items asked in questionnaires 2 and 3 justified the construction of three conceptual factors. The first factor was the valence of the activity tracker, which was formed by the following 12 items: usefulness (practicality), niceness, modernity, amusingness, credibility, ease of use, level of answering to needs, beauty, robustness, intrusiveness, embarrassment, and nuisance. This factor corresponds with the hedonic quality in (Hassenzahl, 2003). The second factor was the preciseness of the activity tracker, which was formed by the following 4 items: exactness, level of detail, clarity, and credibility. This factor corresponds with the pragmatic qualities in (Hassenzahl, 2003). The third and final factor was perceived efficacy of the activity tracker, which was constructed by averaging 3 items on perceived efficacy of the tracker, namely activity increase, health changes, and well-being. This third factor does not correspond with hedonic nor pragmatic qualities but has to do with the perceived effectiveness of the activity tracker. The results of the principal component analysis are reported in Appendix 3.

On the basis of our analysis, some items were left out of our analysis; see Appendix 2 under “left out” for each questionnaire. Four items, of which the validity could not be determined satisfactorily, were nevertheless included in the analysis. One item concerned technological aptitude, and three further items concerned the psychological traits: rebellion, independence, and health-mindedness. We advise to treat the results of these items with caution.

**Statistical Analysis**
We determined the relative importance of all included predictors by using Random Forest, an analysis technique based on recursive partitioning (Breiman, 2001;
Strobl et al., 2009). Random Forest is an ensemble method that makes use of a large number of decision trees, strengthened by “bootstrap aggregating”: drawing random samples from the original dataset with replacement. For each of the bootstrap samples that are drawn, a decision tree is constructed. At each branch of the tree, a random selection of the predictor variables is considered. The variable that produces the best split (ie, most informative and offering the largest contrast) is used to divide the cases over two daughter nodes.

To predict the outcome variable for a specific case, Random Forest uses the predictions of all trees to arrive at an “ensemble” prediction (in the case of regression, it averages the prediction of all trees). To evaluate the performance of the Random Forest model, we can test each tree on those cases that fell outside its bootstrapped sample and thus, were not used to grow the tree. This produces an “out-of-bag” error rate, which is a good approximation of the test error.

Random Forest analysis can produce a list of predictors, sorted by relative importance. This is done by calculating the mean squared error (MSE) and looking at the relative increase of the MSE when the values of a predictor are permuted across cases. Permuting the predictors retains frequency information but destroys the association between the predictor and the outcome variable. If the variable is important for the Random Forest model, we would expect its predictions to
deteriorate and the MSE to go up. Thus, the relative increase in MSE is used to
determine an importance ranking of predictors, sorted from greatest to least
increase in MSE.

Random Forest modeling has the benefit of being able to deal with large
numbers of predictor variables with complex interactions, especially in situa-
tions with relatively few cases relative to the number of predictors. Furthermore,
Random Forest is capable of detecting nonlinear relations between independent
and dependent variables. Random Forest analysis methods have recently been
applied successfully in genetics, clinical medicine, bioinformatics, and the social
sciences (see Strobl et al., 2009; Walsh, Ribeiro, & Franklin, 2017 for examples).

The predictor variables for the Random Forest analysis were taken from the
responses to the questionnaires. All parameter settings, source code, and data files
for the Random Forest analysis are available through the Open Science Foundation.

We used R 3.3 for analysis (R Core Team, 2017) and the R package randomForest
(Liaw & Wiener, 2002) for Random Forest modeling.

Results

Fitbit Use
The mean number of days that participants used their Fitbits was 129.3 (nonconsec-
utive) days ($SD = 88.5; median = 122$). Figure 3.1 shows the distribution of total days
of use.

A graphical overview of usage over time is shown in Figure 3.2. As some users
only started to wear the Fitbit after weeks or even months, the number on the x-axis
refers to the number of days since the first day the device was used, rather than
from the start of the study. The figure indicates both the percentage of participants
who used the device for any length of time after the indicated day and habitual use
(defined as 3, 5, and 7 days worn out of the last 7 days). The decline during the first
50 days coincides for most users with the French holiday season (July-August). The
peak shortly after 100 days coincides with Q2 being sent out (again, for most users).

The pattern of (nonhabitual) usage decline is roughly linear. A linear regression
with time as the independent variable shows a decline of 2.0 percentage points per
week from day 1 to day 300. In other words, every week, 2% (14) of the participants
at start stopped using the tracker entirely. As the base of users is shrinking, this
means that the proportion of participants who stopped using the tracker increases
over time. After 175 days (5.7 months), 50% of users have stopped wearing their
tracker.

Habitual use seems to follow a pattern of slow exponential decay. An exponen-
tial model shows that the proportion of users wearing the activity tracker 5 or

51
more days per week declines 5.7% per week, as calculated from the peak after the summer holiday dip (from day 102 to day 300).

On average, participants took 7492 steps (SD = 3012) per synced day (i.e. day on which they wore their tracker). The median number of steps was 7107, with an interquartile range of 3462. A plot of the distribution of the number of steps taken is provided in Figure 3.3.

An overview of the correlations between the number of days on which the tracker was used, mean number of steps, and a range of self-report measures on personal health are displayed in Table 3.2. The number of days the activity tracker was used significantly predicted mean steps per day: $b = 9.43$, $t(709) = 32.60$, and $p < .001$. A significant proportion of the variance was explained: $R^2 = .08$, $F(1,709) = 8.92$, and $p < .001$. An exploratory analysis of correlations with measures on personal health showed generally weak to negligible associations for both days used and mean steps per day. Strongest associations were between the number

![Figure 3.2: Usage decline over time. The horizontal axis shows the number of days since the first day of use. The percentage of participants who used the activity tracker for any number of days after a particular day is indicated with a solid line. The other lines indicate habitual use: the percentage of participants who used the tracker for at least 3, 5, and 7 days in the preceding 7 days. Note that this includes participants who stop using the tracker and later start using it again. The early dip in use is due to the summer holiday.](image-url)
of days used and self-reported general health ($r = .16$) and between days used and physical shape ($r = .12$).

**Reasons for No Longer Using the Tracker**

In both Q2 and Q3, participants were asked how many of the last 30 days they wore their Fitbit. If the answer was “fewer than 5,” they were asked additionally why they did not wear their tracker (more often) in an open-ended question. The responses were categorized and can be found in Table 3.3.

The results from Q2 (sent after 98 days) indicated that 40 participants (6.3% of all respondents) used the Fitbit fewer than 5 days in the last month. The primary reason for not using the device, given by half of those indicating low or non-use, was technical failure or other technical problems, including empty batteries. Other reasons included losing the device or being on a holiday. Technical problems were also the main reason given in Q3 (sent after 232 days), with 17.5% of all respondents reporting this issue.
Factors associated with usage

We used the Random Forest method to investigate which predictors are associated with continued use of the activity tracker. The total number of days on which the device was worn was used as the outcome variable. As many participants did not respond to all three questionnaires, with those who stopped using their Fitbit less likely to fill in questionnaires 2 and 3, we decided to construct two different models, corresponding to two different groups of participants: 1) participants completing Q1 and 2) participants completing all three questionnaires (Q1 to Q3). For the latter model, we analysed only those participants who did not state technical malfunction of any kind as a reason to quit.

In the first model, the data from those 586 participants who completed Q1, gave permission to use their tracker data, and stated neither technical issues nor lost trackers as a reason to no longer track, were entered. Some predictors were adjusted or recalculated. A complete overview of all questionnaires, the exact questions, the response scales, and any recalculations or adjustments is available in Appendix 2.

Figure 3.4 shows the relative impact of each predictor variable in Q1 on the amount of variance explained, expressed as the relative increase in MSE when the
### Table 3.3
Reasons for not wearing the Fitbit.

<table>
<thead>
<tr>
<th>Reason to not wear</th>
<th>Q2 Count</th>
<th>% of reasons</th>
<th>% of n</th>
<th>Q3 Count</th>
<th>% of reasons</th>
<th>% of n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical failure or difficulty</td>
<td>20</td>
<td>50.0</td>
<td>3.1</td>
<td>106</td>
<td>56.7</td>
<td>17.5</td>
</tr>
<tr>
<td>Lost the device</td>
<td>9</td>
<td>22.5</td>
<td>1.4</td>
<td>24</td>
<td>12.8</td>
<td>4.0</td>
</tr>
<tr>
<td>Forgot to wear</td>
<td>2</td>
<td>5.0</td>
<td>0.3</td>
<td>24</td>
<td>12.8</td>
<td>4.0</td>
</tr>
<tr>
<td>No use for device / no motivation</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>16</td>
<td>8.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Health issues</td>
<td>1</td>
<td>2.5</td>
<td>0.2</td>
<td>7</td>
<td>3.7</td>
<td>1.2</td>
</tr>
<tr>
<td>Used other device</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>2</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Was on holiday</td>
<td>7</td>
<td>17.5</td>
<td>1.1</td>
<td>2</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Did not start yet</td>
<td>1</td>
<td>2.5</td>
<td>0.2</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>6</td>
<td>3.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>100</td>
<td>6.3</td>
<td>187</td>
<td>100</td>
<td>30.8</td>
</tr>
</tbody>
</table>

**Figure 3.4.** Plot of relative importance of predictors of sustained use in questionnaire 1 (Q1); BMI: body mass index, MSE: mean squared error.
Chapter 3

predictor is randomly permuted across participants. The Random Forest model with all Q1 predictors included explains 8.29% of variance. Only those predictors whose increase in MSE is above zero are displayed because decreases can be safely attributed to noise.

We explored the effect of the different predictors on the number of days the participants wore their tracker. Figure 3.5–3.9 provide boxplot representations of the distributions for the marginal means (with all other factors kept constant) of the different levels of each predictor. Only those predictors of which the differences in marginal means implies a meaningful difference in real life (> 1 day) are included: Age, goal to quit smoking, iPhone type, sports activities in the company of others, household type, household size, having a smartphone, having an iOS-based-smartphone, profession, and smoking.

Figure 3.5, boxplots of the distributions of Age levels.
The older the participant, the longer the sustained use, apart from > 65yr.
Determinants for Sustained Use of an Activity Tracker: An Observational Study

Figure 3.6, boxplots of the distributions of iPhone type levels. Holders of iPhones show less sustained use than those of other smartphones.

Figure 3.7, boxplots of the distributions of Having the goal to quit smoking. Those not wanting to quit smoking (including non-smokers) have longer sustained use than those who do.
Figure 3.8, boxplots of the distributions of Household type. Single parents show shorter sustained use than other household types.

Figure 3.9, boxplots of the distributions of Sports in company of others. Those who practice individual sports or with relatives, have longer sustained use of the tracker than those who participate in sports with friends or acquaintances.
In the second model, data from 397 participants who completed all three questionnaires (Q1 to Q3) and who did not state technological malfunction as a reason to stop tracking, were entered. Once again, some predictors were adjusted or recalculated. A complete overview of all questionnaires, the exact questions, the response scales, and any recalculations or adjustments is available in Appendix 2.

Figure 3.10 shows the relative impact of each predictor variable in the questionnaires on the amount of variance explained, expressed as the relative increase in MSE when the predictor is randomly permuted across participants. The total percentage of variance explained by the Random Forest model with all Q1 to Q3 predictors is 10.91%.
Once again, we explored the effect of the different predictors on the number of days participants wore their tracker. Figures 3.11–3.15 show the marginal means (with all other factors kept constant) of the different levels of each predictor. Again, only those predictors of which the differences in marginal means implies a meaningful difference in real life (> 1 day) are included: age, perceived effect on goals, user experience (valence), user experience (effect), perceived effect on goals, iPhone type, and the goal to change eating habits. User experience (valence and effect) and perceived effect on goals are continuous variables, so no levels of marginal means could be shown. Instead, we show a partial dependence plot. The user experience variables are shown normalized with a mean 0 and a SD of 1.

Figure 3.11, boxplots of the distributions of the marginal means for participant age.
The under-25 use the Fitbit for a shorter duration than the other groups.
Determinants for Sustained Use of an Activity Tracker: An Observational Study

Figure 3.12, boxplots of the distributions of having the goal to change eating habits. The stronger the goal, the less sustained use of the tracker.

Figure 3.13, plot of the partial dependence of sustained tracker use on the perceived effect of the tracker on goal attainment. A larger perceived effect leads to longer sustained use.
Figure 3.14, plot of the partial dependence of sustained tracker use on user experience of the valence of the tracker. A better user experience leads to longer sustained use.

Figure 3.15, plot of the partial dependence of sustained tracker use on user experience of the efficacy of the tracker. A better user experience leads to longer sustained use.
Discussion

Principal Findings
This study examined the use of an activity tracker, as well as reasons to stop using the tracker and predictors of sustained use. This study shows that of the 711 initial participants, approximately 50% still used their tracker after 6 months, and 12% continued to use their tracker even after 300 days. This rate of decay in usage confirms earlier findings (Thorup Msn et al., 2016). This result also confirms the notion that wearable activity trackers are not subject to the rapid exponential decay of use we see in mobile phone apps, where usually 80% of users drop out in the first few days (e.g. in Chen, 2015).

Reasons to Quit Tracking
After half a year, half of the participants quit using their tracker; at the sending out of Q3, three-quarters of participants were no longer using the Fitbit. When asked for reasons for their quitting, 56.7% of those who answered the question stated some form of technical malfunctioning (including empty batteries). A further 12.8% indicated they had lost the device. This result confirms findings (e.g. Harrison et al., 2014) in which trackers were abandoned because of empty batteries, with the perceived cost of replacement too high or too cumbersome (Ledger & McCaffrey, 2014), but this result deviates from other findings (e.g. Clawson et al., 2015; Harrison et al., 2015) in which technological failures comprise only a very small part of reasons to no longer use a device, and reasons such as sustained motivation, device aesthetics, and device accuracy seem to be of more importance. However, in studies in which reasons to quit tracking are covered, either the demand characteristics of the study (e.g. Harrison et al., 2015) or the data gathering technique (e.g. Clawson et al., 2015, where advertisement data from Craigslist were used) preclude the reliable registration of technical failure as a reason to quit tracking. The results from this study may therefore serve as a first indication of the relative importance of technological reliability for the sustained use of feedback technology. Simple actions, such as changing a battery, already seem to raise insurmountable barriers for sustained use of a tracker. Newer activity trackers, fortunately, mostly do not rely on button cell batteries, which take some effort to replace, but make it possible to recharge the device, much like one would recharge a smartphone. This, however, also constitutes a barrier to sustained use. Further research into the effect of technological failures on sustained use of activity trackers, and how to help users overcome the barriers brought about by technical issues such as empty barriers, is needed to corroborate this finding and shed light on potential solution strategies.
Relative Importance of Predictors for Self-Tracking

In the analysis of items from Q1 only (Figure 3.4), participant age was the only predictor that showed great impact; higher age was associated with more sustained use. A second group of predictors for sustained use were the goal to quit smoking (with not having the goal associated with longer use), iPhone type (with not having an iPhone associated with longer use), and household type (with single parents using the tracker for a shorter duration than all other groups). All other predictors were not found to have a noteworthy impact on sustained tracker use.

In the analysis of all (three) questionnaires (Figure 3.10), participant age was once again the strongest predictor of sustained tracker use. As in Q1, sustained use increased with age (to a point), overall (Q1 to Q3) the Fitbit was used for a shorter duration by the youngest age group (under 25). Other important predictors were user experience–related predictors (tracker valence and user experience, and perceived efficacy and helpfulness of the tracker), iPhone type, having the goal to change eating habits, and wanting to quit smoking were other relevant predictors, albeit in an opposing way: these predictors were associated with a decreased use of the self-tracking device.

The relative importance of age as a predictor of sustained tracker use (with higher age associated with longer use) is in line with literature (e.g. Bossen et al., 2013; Couper et al., 2010; Davies et al., 2012). More research is needed to answer the question why this is, and to determine its implications for the design of tracker-based interventions for physical activity. Is the greater efficacy for older participants problematic, or are younger participants already well-served by other possibilities to exert themselves physically? In a similar notion, is this age-effect a consequence of self-selection, in which younger people already have enough alternatives for physical activity, and it is mostly those older than 25 years that turn to tracking as viable solution? Answers to these questions also have implications for the development of tracker interventions. Do we need more age-inclusive solutions, or can we regard this type of intervention as more effective for people older than 25 years?

The importance of user experience–related predictors such as valence, perceived efficacy, and preciseness of the tracker is also in line with previous studies. User experience and ease of use (Clawson et al., 2015; Lazar et al., 2015), functionality or lack thereof (Clawson et al., 2015), the possibility to upgrade toward a newer device (ibidem), aesthetics and form (Harrison et al., 2015), and perceived fit between device and self-image (Lazar et al., 2015) have all been cited as reasons to either abandon the tracker or to keep using it. This sheds light on the relative importance of technological and design-related aspects of behaviour change feedback technology aimed at greater physical activity. Even though there is a substantive literature on the subject, this remains an underexposed area in current health
behaviour change research. Clunky intervention designs carry the risk of being rejected by their participants and, more importantly, a lack of uptake once the intervention hits the market or the app store. In health behaviour change research, a lot more attention is needed for user experience, user friendliness, and the aesthetic experience.

The negative impact of goals, such as wanting to change one’s eating habits or wanting to quit smoking, seem logical in hindsight. The Fitbit tracker does not in itself contribute to the attainment of these goals, which could easily have a demoralizing effect. The fact that having an iPhone seemed to reduce the chances of sustained use could point at another covert measure of user experience. The iPhone interface for Fitbit-related feedback might possibly be more difficult to use or less functional than its Android equivalent; alternatively, iPhone users may be psychographically different from Android users on traits that lead to reduced usage of activity trackers. However, no evidence to support either hypothesis is presently available.

A surprising finding was the lack of effect of a range of predictors, which is not in line with literature (e.g. Fischer, 2008; Hermsen, Frost, Renes, et al., 2016). Socioeconomic status markers such as education (e.g. Couper et al., 2010; Geraghty et al., 2013) and profession (e.g. Al Asadi et al., 2014; Habibović et al., 2014), gender (e.g. Couper et al., 2010; Geraghty et al., 2013; Wanner et al., 2010), psychological traits (Baumeister & Heatherton, 1996; Braverman, 2008; Kuhl, 1985), personal health-related factors (e.g. Grembowski et al., 1993; Strecher et al., 1986), strong motivation (e.g. in Bossen et al., 2013; Davies et al., 2012), strong, clear goals (e.g. in Couper et al., 2010), and social interaction (e.g. in DiMatteo, 2004) did not appear to affect tracker use. These have often been researched out of context, with predictor singled out and assessed independently. The current result could point to the fact that some predictors may not be as important as we think they are, when compared with many other possibilities. When placed in context, their role may be smaller than we assumed. A competing hypothesis, however, could be preselection; for instance, it is possible that motivation did not play a large role, because those who entered the challenge were already highly motivated. Similarly, perhaps only those already high on psychological traits such as conscientiousness took part. This preselection would limit the confidence in some of the null-results found in this study. If so, however, this preselection constitutes less of a problem as one would think. We can assume similar preselection would take place in the market place; it is reasonable to suspect that traits and states found in those who take part in this study would resemble states and traits of those people who would be interested in using a Fitbit in the first place. Unfortunately, this cannot be deducted from our research. Further research would be interesting.
The entire range of independent variables in Q1 explained 8.29% of variance; in Q1-Q3, the whole set of predictors accounted for 10.91% of explained variance. In Cohen’s (1992) frequently used assessments of effect sizes for psychology, an $R^2$ of .095 (Q1) to .099 (Q1–Q3) is described as a small effect or approaches a medium effect. Such an effect size is common in social and behavioural sciences, for situations where there is a lot of individual variation and many different factors may affect the dependent variable independently (see also Abelson, 1985). To our best knowledge, this is the first quantitative study looking into the factors influencing the persistent use of activity trackers. Earlier studies (e.g. Fritz et al., 2014; Harrison et al., 2014; Munson & Consolvo, 2012; Rapp & Cena, 2015; Rooksby et al., 2014; Shih et al., 2015) were qualitative and small-scale studies (7 to 31 participants) and did not attempt to model activity tracker use. Thus, we have no immediate context to compare our model’s performance with.

Intuitively, we may have expected a larger effect size from such a broad range of predictors. We can discern two competing hypotheses. A first hypothesis is that sustained use is mostly predicted by random events such as empty batteries or loss, but there are many small but significant contributions from a broad range of predictors. A second hypothesis is that unmeasured third variables are responsible for the relative lack of effect. Not all relevant predictors we could identify in the literature were included in the questionnaires. First, perceived self-efficacy was not directly assessed but only through measures regarding perceived efficacy of healthier behaviour change. Second, literature (Rooksby et al., 2014) suggests that different tracking styles exist, such as tracking physical activity to diagnose a secondary problem such as sleeping disorders or stomach problems, or “fetishized” tracking: tracking because it is cool or otherwise desirable. In this study, the tacit assumption is that all participants want to at least document and probably also change their physical activity, which might not be the case in reality. Third, different forms of intrinsic motivation, such as motivation for autonomy, mastery, and relatedness (Deci & Ryan, 2000), might lead to different levels of adherence to activity tracking. Finally, the completion of set goals was not registered. It is plausible to assume that when people achieve their goal, their interest in tracking their progress wanes. The inclusion of these possible moderators in future research would shed light on their effect on sustained use of a tracking device. Further research could shed light on which of these possible explanations would be most feasible.

Limiting Factors
A few limitations to the current study warrant further discussion. Firstly, our confidence in the validity of the findings is limited by the fact that of the original 929 participants, 711 gave permission to access their Fitbit data and filled out Q1;
of those 711, only 575 took part in Q2 (80.9%), and 542 took part in both Q2 and Q3 (76.2%). The greater part of those participants who did not fill out Q2 or Q3 quit using their Fitbit somewhere in the period preceding that questionnaire. Even though this decline in adherence is not at all uncommon in interventions for health behaviour change, and thereby no cause for alarm, their data would have increased the validity and reliability of our findings.

Similarly, 56.7% of those who provided a reason for their no longer tracking stated technical malfunctioning. Of those who did not report a reason (e.g. because they did not fill out Q2 or Q3), we do not know why they no longer took part. However, the fact that at least 17.5% of all participants quit because of technical reasons still emphasizes the importance of this finding, regardless of the reasons the non-reporters could have had for quitting.

A second, and possibly greater, limitation to the validity of the findings stems from the way the study design was carried out. Questionnaire construction and data gathering were carried out by the MySantéMobile team. The quality of the questionnaires would have benefited from early involvement of social scientists with relevant experience in questionnaire construction, which would have led to a more hypothesis-based selection of questions, and more informative response scales. As it is, we think this study has enough validity to serve its purpose, that is, as an exploratory analysis of potential determinants of sustained use of physical activity trackers.

A third limitation in the study design is the fact that the psychological predictors of use such as the Big Five and user experience-related predictors were not included in Q1 but made a first appearance in Q3. This limits the applicability of findings concerning these predictors because only participants making it to Q3 (76.2% of those who filled out Q1 and gave access to their data) answered these questions. However, psychological traits are known to be stable (Cobb-Clark & Schurer, 2012), so it is reasonable to expect that no great changes in big five traits occurred. User experience-related predictors can only be measured once participants have used the product; an a-priori judgment lacks value. These, therefore, could not have been included in Q1.

Finally, the data analysis method selected has its benefits, such as robustness toward overfitting and good handling of relatively low participant populations, but Random Forest analysis also has its limitations. The result of the analysis is a ranking of the relative importance of each predictor on the use of the activity tracker. Due to its ensemble nature, results from a Random Forest can be hard to interpret (unlike a linear model). Contributions from a variable can be present in multiple ways and through nonlinear and/or (higher-order) interactions. However, through Random Forest modelling, we can establish which predictor variables
are important with respect to outcome variables. These variables can be studied further to establish their effect and interactions with other variables.

Conclusions

This study confirms earlier findings that habitual use of an activity tracker tends to decline at a slow exponential pace rather than show the rapid exponential decline shown in health app use. When they start using an activity tracker, most users in our sample continued to use it for at least half a year. Around 12% of users still use their tracker after 300 days.

This study also shows that sustained use of an activity tracker is not easy to predict. Most known predictors of sustained adherence to physical activity interventions do not seem to have an impact on sustained use in the sample observed in this study. When participants no longer use their tracker, technological failures such as empty batteries seem the predominant reason to quit.

The broad range of predictors entered in the Random Forest model in this study only led to a small proportion of explained variance. Those predictors that did have an effect on sustained use were participant age and factors related to the user experience of tracker use.

Regardless of the limitations to the findings cited above, this study shows some much-needed insight in predictors of sustained use of trackers. Furthermore, this study is one of few examples in which academia gets the chance to evaluate data from industry; the field would greatly benefit from a greater number of such collaborations, preferably with a larger role for the academic partner in setting up the study.
Chapter 4: Evaluation of a Smart Fork to Decelerate Eating Rate

Abstract

Eating rate is a basic determinant of appetite regulation, as people who eat more slowly feel sated earlier and eat less. Unfortunately, without assistance, eating rate is difficult to modify due to its highly automatic nature. The 10SFork, designed by Slow Control, Paris, provides feedback to raise awareness of eating rate in order to help people eat more slowly. It records behaviour and provides real-time haptic feedback on individual eating rates.

11 participants (3 male, 8 female, \( M(\text{age}) = 21.2 \)) used the fork both in a laboratory setting and at home. All participants indicated having high eating rates. We interviewed them on perceived efficacy, acceptability, comfort, accuracy, motivation, and sustained use of the fork. Participants feel the 10SFork is an acceptable tool to decelerate eating rate. Participants were more aware of their eating rate, but this did not always lead to behaviour change. The fork is generally seen as comfortable and sufficiently accurate. The vibrotactile feedback worked as expected, but the visual feedback largely remained unnoticed. Sustained motivation to use the fork was limited because participants did not see themselves as the product’s target group.

---

Introduction

Overweight is associated with a range of negative health consequences, such as type II diabetes, cardiovascular disease, gastro-intestinal disorders, and premature mortality (Berenson, 2012). One promising means to combat overweight is through encouraging people to eat more slowly (Martin et al., 2007). People who eat quickly tend to consume more (De Graaf & Kok, 2010; Robinson, Thomas, Aveyard, & Higgs, 2014; Viskaal-van Dongen, Kok, & De Graaf, 2011) and have a higher body mass index (Lee et al., 2011; Maruyama et al., 2008; Otsuka, Tamakoshi, & Yatsuya, 2006; Tanihara et al., 2011) while people who eat more slowly feel sated earlier and eat less (Cassady et al., 2009; Kokkinos et al., 2010; Rolls, 2007; Zijlstra et al., 2009). Unfortunately, eating rate is difficult to modify, due to its highly automatic nature (Petty, Melanson, & Greene, 2013). In clinical settings, researchers have had some success changing behaviour using devices that deliver feedback in real time (Bergh et al., 2008; Ford et al., 2009; Van Elburg et al., 2012). However, existing technologies are either too cumbersome (Zandian et al., 2009) or not engaging enough (Hamilton-Shield et al., 2014) for use in daily life contexts. Training people to eat more slowly in everyday eating contexts, therefore, requires creative and engaging solutions. The purpose of this chapter is to present a qualitative evaluation of the feasibility of a smart fork to decelerate eating rate in daily life contexts. Furthermore, we outline the planned research to test the efficacy of this device in both laboratory and community settings.

Evaluation

Assessment

We performed a qualitative study to assess the acceptability, perceived efficacy and user experience of the SmartFork. The augmented fork contains sensors and actuators that provide real-time feedback (see Figure 4.1). The fork delivers feedback at 10-second intervals between bites. If users take a bite too quickly (i.e. before the end of the 10-second interval), they feel a gentle vibration in the handle of the fork and see a red indicator light.

The fork provides a series of data recording methods. First, the fork determines the exact time at which the meal is started and ended (i.e. meal duration). Second, it counts the total number of bites per meal and per minute (i.e. eating speed). Third, it calculates the average interval between bites and, fourth, determines the ratio of over-speed bites. The fork stores all data for later review via USB or Bluetooth. The desired interval between bites and feedback modalities (lights and vibrations) can
be adjusted in an online control panel. In addition to the vibrotactile and visual feedback, the fork is connected to a secure online platform. After logging on to the platform, users can review their past behaviour: number of bites, percentage of bites eaten too quickly, and duration of the meals. Possibilities for sharing and integration with social media are provided.

To test this fork, 11 participants (3 male, 9 female, age 18–35, all self-perceived fast eaters ($m = 7.2$, $SD = 1.82$ on a scale from 1 to 10, where 1 is ‘extremely slow’ and 10 is ‘extremely fast’) ate a meal using the fork in our laboratory. Subsequently they used the fork for three consecutive days in their home setting, eating as many meals as possible with the fork. All participants ate the main meal of the day, dinner, with the fork. Three participants also used the fork for other meals including breakfast and lunch. After the laboratory meal and upon returning the fork, participants shared their experiences in semi-structured interviews covering the following topics: perceived effect on eating rate, comfort of use, feedback accuracy, social aspects of fork use, and motivation for using the fork. Interviews were recorded and transcribed, and a thematic classification on the transcripts was...
performed. The study protocol was approved by the Institutional Review Board of the Faculty of Social Sciences of <blinded for review>. All participants provided written informed consent.

**Participant feedback**

All participants felt that the feedback was generally accurate and consistent and found the technology acceptable. Everyone found the fork’s size and weight acceptable, felt the fork was easy to handle, and felt that the fork’s vibrotactile feedback was not uncomfortable, but could not be ignored either. While each participant reported some false positives, e.g. vibrations when not taking a bite, no participant saw that as a threat to the usability of the fork. However, all participants found it hard to estimate when the ten-second wait was over. Interviews suggest the fork may result in changes in both perceptions and behaviour. All participants report a heightened awareness of eating rate and all but one participant reported that they ate more slowly when using the fork. When eating in company, none of the participants felt ashamed when using the fork; rather, it sparked humour and started some lively conversations about eating rate and healthy eating. Surprisingly, a few participants reported some frustration with decelerated eating rate, expressing a desire to return to their former speedier eating habits.

All participants were motivated to try the fork. After a few meals, however, motivation waned in a minority of the participants; the majority remained motivated to use the fork throughout the three-day period. All participants could imagine the fork being effective in retraining eating rate in the long run. Yet, none of the participants felt they were part of the product target group, i.e. they did not perceive their high eating rate as a major problem for their health.

**Conclusions**

The 10sFork has the potential to become a successful intervention in slowing down eating rate. Users feel it is an acceptable product that is sufficiently comfortable and accurate. They report enhanced awareness of their eating rate and feel comfortable using the fork in social settings. However, self-perceived target group membership, and the incapacity of the fork to take meal characteristics into account, may be issues affecting acceptance of the fork as an intervention for healthy eating in real life.

To formally evaluate the efficacy of the 10sFork in slowing down eating rate, we have received funding of the Netherlands Organisation of Scientific Research (NWO). We will conduct two studies. The first study will assess the effect of the
Chapter 4

feedback on eating rate, satiety, and intake in a single, standardized meal. In the second study, we will examine the efficacy of the fork over time in naturalistic eating contexts. Results from these studies will contribute to answering the question of whether this tool can be a viable instrument to reduce eating rate, and control food intake.
Evaluation of a Smart Fork to Decelerate Eating Rate
Chapter 5: The effect of real-time vibrotactile feedback on eating rate, satiation, and energy intake

Abstract

Eating rate is a basic determinant of appetite regulation, as people who eat more slowly feel sated earlier and eat less. Without assistance, eating rate is difficult to modify due to its automatic nature. In the current study, participants used an augmented fork that aimed to decelerate their rate of eating (i.e., bite frequency). A total of 114 participants were randomly assigned to the Feedback Condition (FC), in which they received vibrotactile feedback from their fork when eating too fast, or a Non-Feedback Condition (NFC) in which they did not receive feedback. Results indicated that participants in FC ate at a slower eating rate, or took less bites per minute, than those in NFC, \( t(101.63) = 2.58, p = 0.011, d = 0.52 \). This, however, did not lead to significant changes in the amounts consumed, \( t(100.92) = -0.26, p = 0.80 \), or levels of satiation, \( t(96.4) = 0.24, p = 0.81 \). In addition, it was found that participants in FC consumed fewer grams per minute, \( t(101.54) = 2.10, p = 0.038, d = 0.43 \), but not per bite, \( t(101.27) = 0.54, p = 0.59 \).

These findings indicate that vibrotactile feedback delivered through an augmented fork is capable of reducing eating rate but does not directly influence levels of satiation or overall food consumption. Overall, this study shows that vibrotactile feedback can be a viable tool in interventions that aim to reduce eating rate. The long-term effectiveness of this form of feedback on satiation and food consumption, however, awaits further investigation.


* shared first authorship
The effect of real-time vibrotactile feedback on eating rate, satiation, and energy intake

Introduction

The worldwide prevalence of overweight and obesity are cause for concern (Finucane et al., 2011). A promising means to combat overweight may lie in reducing eating rate (Martin et al., 2007; Robinson et al., 2014). People who eat quickly tend to consume more than slower eaters (De Graaf & Kok, 2010; Robinson et al., 2014; Viskaal-Van Dongen et al., 2011) and feel less sated after a meal (Rolls, 2007; Zijlstra et al., 2009). Moreover, there is a cross-sectional association between eating rate and obesity; people who eat at a faster rate are more likely to be overweight or obese (Ohkuma et al., 2015; Otsuka et al., 2006; Tanihara et al., 2011).

Eating rate may influence satiation levels and energy intake through a number of mechanisms. When people eat slowly, this influences the secretion of satiety hormones such as insulin and glucagon-like peptide 1 (Cassady et al., 2009; Kokkinos et al., 2010). Slower eating also increases food oral exposure (Bolhuis et al., 2011; Weijzen, Smeets, & De Graaf, 2009) and the number of chews per unit of food (Bolhuis et al., 2013a, 2013b), which have both been shown to lower energy intake (Bolhuis et al., 2013a, 2013b; Weijzen et al., 2009). Finally, slower eating may decrease feelings of deprivation by enhancing and prolonging pleasurable aspects of eating (Brownwell, 2000).

One barrier to changing eating rate is that it may be a highly automatic behaviour, making eating rate difficult to change (Wilson, 2002). However, recent research suggests that real-time feedback can interrupt the execution of deeply engrained habitual behaviours and make them available for conscious scrutiny and behaviour change (Hermsen, Frost, Renes et al., 2016). Furthermore, feedback is known to have motivational consequences, giving higher priority to the behaviour that is the target of the feedback (Northcraft et al., 2011).

In the case of eating rate, visual and auditory mealtime feedback has been used to give eaters feedback on how much and at what rate to eat during a meal (Zandian et al., 2009). This method has been found to be effective in reducing food intake and promoting weight loss, both in clinical as well as non-clinical contexts (Ford et al., 2009; Ioakimidis, Zandian, Bergh, & Södersten, 2009; Zandian et al., 2009). A potential limitation of this type of feedback, however, could be that it can be too cumbersome or artificial to use in real-life eating contexts. Real-time vibrotactile feedback, the presentation of simple vibrations as a means of conveying alerts or information (Hoggan et al., 2009; Qian et al., 2011) may present a viable alternative to visual and auditory mealtime feedback on eating rate. Vibrotactile feedback can provide straightforward real-time signals with little disruption to the visual and auditory channels (Hale & Stanney, 2004; Sigrist, Rauter, Riener, & Wolf, 2013). This form of feedback has been shown to improve motor skill acquisition (Spelmezan, Jacobs, Hilgers, & Borchers, 2009; Van Erp, Saturday, & Jansen, 2006),
rehabilitation and posture control (Alahakone & Senanayake, 2009, 2010), and navigation and way finding (Heuten, Henze, Boll, & Pielot, 2008; Van Erp & Van Veen, 2004). Real-time feedback may also raise awareness about one’s speed of eating without interrupting conversations or other pleasurable aspects of a meal. By doing so, this method may be more easily applied to reduce people’s eating rate within real-world eating environments. However, little is known about the utility of real-time vibrotactile feedback to modify eating rate.

This study therefore set out to assess the effect of real-time vibrotactile feedback on eating rate, satiation, and ad-libitum food intake. In the present study, we used an augmented fork that contains sensors and actuators that provides people with vibrotactile feedback when they are eating too fast. Specifically, the fork delivers real-time feedback at 10-second intervals between bites. If users take a bite too quickly (i.e. before the end of the 10-second interval), they feel a gentle vibration in the handle of the fork. Although previous research indicates that the fork is perceived as a comfortable, accurate, and effective method to decelerate eating rate (Hermsen, Frost, Robinson, et al., 2016), it is still unclear whether vibrotactile feedback affects users’ subsequent eating behaviour. To examine this question, we conducted an experiment in which the real-time vibrotactile feedback of the fork was manipulated (i.e. vibrotactile feedback versus no feedback). First, we hypothesized that participants who received real-time vibrotactile feedback would decelerate their eating rate, conceptualized as eating fewer bites per minute and eating more bites outside the designated 10-second time interval, compared to those who did not receive feedback. Second, we hypothesized that changes in eating rate would lead to increased satiation and decreased ad-libitum food consumption.

Materials and Methods

Experimental design and stimulus materials
An experimental design with a single between-subjects factor (vibrotactile feedback versus no-vibrotactile feedback) was used. To provide participants with real-time feedback while eating, we used the 10sFork (SlowControl, Paris, France, see Figure 4.1). This fork contains sensors to measure eating rate and actuators to deliver vibrotactile feedback when the user eats too quickly. In the Feedback Condition (FC), participants ate a lunch meal with the augmented fork. If participants took a bite too quickly (i.e. before the end of a pre-set 10 second time interval between bites), they felt a gentle vibration in the handle of the fork and saw a red indicator light. Pre-tests showed that this 10s bite speed slows down fast eaters, without making it too difficult for them to finish their meal (Hermsen, Frost, Robinson, et
The effect of real-time vibrotactile feedback on eating rate, satiation, and energy intake

al., 2016). In the No-Vibrotactile Feedback Condition (NFC), participants ate the same lunch meal with the same augmented fork but did not receive any feedback regarding their eating rate. Participants were randomly assigned to either the FC or NFC condition. The size, weight and design of the augmented fork resembled a normal fork. The present study and its primary and secondary outcome measures were pre-registered in the Dutch Trial Register (NTR5237).

Participants
To be able to detect a medium effect size, with a power of 0.80 and a significance level of 0.05, 64 participants in each experimental condition were required. Therefore, we aimed to recruit 128 participants. Due to practical constraints, the total sample that was recruited consisted of 123 participants, of which 63% were female (n = 77). Participants were mainly undergraduate or graduate students at Radboud University (63%), or non-students, e.g. employees of the university or other institutions and companies (37%). Five participants were excluded before testing because of BMI scores (BMI: kg/m² ≥ 35) that did not comply with our inclusion criteria. Four participants were excluded after testing because their fork data showed severe inconsistencies (e.g., one participant appeared to have consumed 296 grams in only 30 seconds). Therefore, the final sample consisted of 114 participants (70 female, 44 male) (see Figure 5.1 for the CONSORT Flow Diagram). The mean age of participants was 29.05 (SD = 13.16). Participants’ mean BMI was 23.51 kg/m² (SD = 3.36). In our sample, 75% of participants had a normal weight (18 ≥ BMI ≤ 25 kg/m²) and 25% were overweight or obese (25 ≥ BMI ≤ 35 kg/m²).

Procedures
All participants were recruited through an internet sign-up program at the Behavioural Science Institute (BSI) of the Radboud University or via direct approach at campus. Specifically, we asked participants to register for our study if they considered themselves to be a fast eater and were motivated to learn to eat slower. The study was described as an investigation of the usability of a smart fork to help people to eat slower. Registration for our study was open to participants between 18 years and 80 years of age who had a BMI between 18 and 35 kg/m². Participants were instructed to refrain from eating for three hours before participation in our study to control for individual variations in hunger. The study and all procedures involved received approval from the Ethics Committee of the Faculty of Social Sciences at Radboud University.

Data collection took place on weekdays between 11.30 AM and 2.30 PM in the period May – December 2015. To simulate a relatively naturalistic eating setting, the experiment took place in a laboratory furnished as a small restaurant (for a
detailed description of this room, see Hermans, Larsen, Herman, & Engels, 2012). All participants sat at single tables, separated by screens to avoid visual contact with the other participants in the room. A maximum of three people participated in one experimental session; if more than one participant took part in one single session; all participants were assigned to the same experimental condition.
The effect of real-time vibrotactile feedback on eating rate, satiation, and energy intake

Participants were asked to read and provide written consent, after which the experimenter measured each participant’s weight and height (Lohman, Roche, & Martorell, 1992). Participants then completed a series of questions to assess their self-perceived eating rate, perceived detrimental effect of their eating rate and any possible conditions that could influence their appetite or the consumption of the meal (e.g. colds, allergies). Then, in order to keep instructions constant over both conditions, all participants were told about the potential positive health effects of eating slowly and the potential of a smart fork to help them to achieve this goal. All participants were told that their fork would monitor their eating rate, but only the participants in the FC were told about the possibility of receiving a gentle vibration in the handle of the fork when eating too fast. After some final instructions on how to switch the fork on or off, participants were then served a lunch meal, consisting of 800 grams of Pasta Bolognese (or vegetarian equivalent; see Table 5.2 for the caloric and macronutrient content of both meals). The lunch was served in a large bowl, from which participants could self-serve their lunch. In this manner, participants could select their own portion size. Furthermore, participants were told that they could eat as much or little as they wanted. The experimenter asked participants to directly switch the fork on/off when starting and finishing their meal, before leaving the room. Participants were not offered any drinks, neither were they allowed to drink their own beverages, during consumption of the meal.

After approximately ten minutes the experimenter checked whether participants had finished their meal. If this was the case, the experimenter collected the uneaten food. No time duration was set for participants to finish their meal. After consuming the meal, participants were asked to complete some post-meal questions about their satiation level, their perceived eating rate during the meal, the effect of the fork on their eating rate, and their overall impression of the study. After the participants had completed this questionnaire, they received a short debriefing about the purpose of the study. Participants received partial course credit or a gift voucher (€7.50) for their participation. After all data were collected, participants were fully debriefed about the study by e-mail.

Measures

Descriptives
BMI. Participants’ weight and height were measured following standardised procedures. Height was measured to the nearest 0.5 cm using a stadiometer (Seca 206; Seca GmbH & Company, Hamburg, Germany) and weight was measured to the nearest 0.1 kg using a digital scale (Seca Bella 840; Seca GmbH & Company). Participants’ BMI was calculated as weight in kilograms divided by height in meters squared. We determined whether participants were underweight, normal weight,
overweight or obese using the International Classification of adult underweight, overweight and obesity according to BMI (WHO, 2017).

Participants' subjective eating rate, perceived detriments and motivation to change (self-report). Participants’ rated how their eating rate compared with other people with one single item on a 5-point scale from 1 (‘very slow’) to 5 (‘very fast’) (before the meal). Furthermore, participants indicated how problematic their eating speed was on a 140 mm VAS scale anchored from 0 ‘not at all’ to 140 ‘very problematic’. Finally, participants indicated their motivation to learn to eat slower on a 140 mm VAS scale anchored from 0 ‘not at all motivated’ to 140 ‘very motivated’.

Manipulation checks

Awareness of eating rate. Participants’ awareness of their eating rate during the experiment was assessed after the meal with two questions. First, participants were asked to indicate how aware they were of their own eating behaviour on a 10-point scale from 1 (‘not at all aware’) to 10 (‘very aware’). Second, they were asked to indicate whether they thought they had consumed their meal at a slower pace than usual. They could answer this question with 1 (‘yes, I ate at a slower pace than normal’), 2 (‘no, I ate a faster pace than normal’), or 3 (‘no, I ate as fast or slow as I usually would do’).

Dependent variables

Primary outcome measures. In both conditions, the 10sFork was set up to automatically record each bite. Based on these data, eating rate (i.e. the total number of bites per minute) and success ratio (i.e. number of bites outside 10-second time interval divided by total bites) were calculated. To measure ad-libitum food intake, a digital scale (Kern 440; Kern & Sohn, Balingen, Germany) was used for measuring amounts served and consumed. At the end of each session, the amount of food consumed in grams was measured. Participants’ total food intake was calculated by subtracting the amounts left on the plate and in the bowl from the initial amount of 800 grams that was served to them.

Secondary outcome measures. Meal duration was calculated as the time in minutes between the first and last bite. These data were recorded by the fork. If participants had not switched off their fork directly after having their last bite, we subtracted the time between last bite taken and the time after which the fork was switched off (n = 4). The total number of fork servings (i.e. number of fork servings during the meal) and average time interval between fork servings (i.e. time in seconds per bite, Hill & McCutcheon, 1984) were also recorded by the fork. Satiation levels were self-reported before and after the meal. Before the meal, participants rated their hunger level on a 140 mm VAS scale anchored from 0 ‘not at all’ to 140 ‘very hungry’ (cf.
Hermans et al., 2013). After the meal, participants rated how satiated they were on the same 140 mm VAS scale anchored from 0 ‘not at all’ to 140 ‘very satiated’.

Post-hoc analyses. In line with other studies on eating rate (e.g. Bolhuis & Keast, 2016), we also conceptualized eating rate as grams of food consumed per minute and average bite size (i.e. amount in grams consumed divided by total number of forks servings). These measures, however, were not included in the original analysis plan that was pre-registered in the Dutch Trial Register.

Statistical analyses
Before testing our hypotheses, we inspected all variables to look for any anomalies. Further, we inspected sampling distributions to test for normality of our data. To detect outliers, two methods were used. First, outliers were identified by visual inspection of the data. In total, we identified seven participants with outliers: two participants showed very long meal durations (> 30 minutes), two participants had a high number of bites (> 90 fork servings) and three participants had very long intervals between bites (> 60 seconds between bites). Second, participants who consistently provided extreme scores (in the most extreme 5%) were noted. This inspection revealed another three participants with extreme scores. Because we decided to exclude these 10 participants from further data analysis, all secondary, primary and post-hoc analyses involved a total of 104 participants. Subsequently, to check for baseline differences, we inspected how strongly potential confounders (i.e. sex, age, BMI, pre-experimental hunger, subjective eating rate, perceived detriments and motivation to change) differed between conditions. We used Cramér’s V to determine whether any of the potential confounders differed with an effect size of moderate strength (cf. Grujters, 2016).

The independent variable was a manipulated, dichotomous variable. All dependent variables in the design were interval variables. Therefore, effect size measure Cohen’s d is an adequate representation of the association between the independent variable (i.e. experimental condition) and independent variables (e.g. eating rate). Effect sizes and their confidence intervals were calculated. Effect sizes of 0.2, 0.5, and 0.8 are indicative of small, medium, and large effects, respectively (Cohen, 1992). All analyses in the present study were performed using the t-test for unequal variances (Ruxton, 2006). To provide additional information about the validity of our statistics, we also report the p values as a secondary measure of significance. In standard analysis, these p values are not corrected for multiple testing. Therefore, we also performed a final analysis in which these p values were corrected for multiple testing. Data were analysed using SPSS for Macintosh version 22 and R: A Language and Environment for Statistical Computing.
Chapter 5

Results

Randomization checks
The conditions did not differ in sex, age, BMI, hunger before meal, subjective eating rate, perceived detriments of eating rate, and motivation to change eating rate, indicating that our randomization procedure was successful (see Table 5.1).

Table 5.1
Randomisation checks

<table>
<thead>
<tr>
<th>Variables measured, by condition</th>
<th>Feedback Condition (FC) (n = 58), M ± SD</th>
<th>No-Feedback Condition (NFC) (n = 56), M ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>27 males, 31 females</td>
<td>17 males, 39 females</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>29.97 ± 14.02</td>
<td>28.08 ± 12.26</td>
</tr>
<tr>
<td>BMI (kg / m²)</td>
<td>24.02 ± 3.20</td>
<td>22.99 ± 3.46</td>
</tr>
<tr>
<td>Hunger before meal</td>
<td>88.68 ± 26.45</td>
<td>96.65 ± 26.09</td>
</tr>
<tr>
<td>Subjective eating rate</td>
<td>3.95 ± 0.51</td>
<td>3.86 ± 0.67</td>
</tr>
<tr>
<td>Perceived detriments of eating rate</td>
<td>37.93 ± 32.75</td>
<td>39.05 ± 30.71</td>
</tr>
<tr>
<td>Motivation to change eating rate</td>
<td>69.83 ± 33.92</td>
<td>67.05 ± 33.22</td>
</tr>
</tbody>
</table>

Table 5.2
Experimental foods used in the study

<table>
<thead>
<tr>
<th></th>
<th>Non-vegetarian meal</th>
<th>Vegetarian meal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice frequency (n)</td>
<td>11 (7 NFC / 4 FC)</td>
<td>103 (40 NFC / 54 FC)</td>
</tr>
<tr>
<td>Energy per 100g (kcal)</td>
<td>202</td>
<td>277</td>
</tr>
<tr>
<td>Fat per 100g (g)</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Protein per 100g (g)</td>
<td>7</td>
<td>15.5</td>
</tr>
<tr>
<td>Carbohydrates per 100g (g)</td>
<td>34.5</td>
<td>30</td>
</tr>
<tr>
<td>Fiber per 100g (g)</td>
<td>3.2</td>
<td>2</td>
</tr>
<tr>
<td>Salt per 100g (g)</td>
<td>1.5</td>
<td>2</td>
</tr>
</tbody>
</table>

1 – NFC: non-feedback condition, FC: feedback condition.

Manipulation checks
Participants in the FC condition did not differ from participants in the NFC in how aware they were of their eating behaviour during the experiment, t (1,102) = -1.31, p = .19. However, participants differed significantly in their self-reported eating rate.
The effect of real-time vibrotactile feedback on eating rate, satiation, and energy intake

during the experiment; participants in the FC reported that they ate more slowly than did participants in the NFC, $t(1,102) = 5.55, p < .001$. Furthermore, participants differed in how much they thought the fork helped them to eat more slowly; participants in the FC had more confidence in the perceived efficacy of the fork to change their eating rate than did those in the NFC, $t(1,102) = -4.40, p < .001$.

**Main findings**

**Primary outcomes**

With regard to participants’ eating rate (i.e. total number of bites per minute), participants in the FC had fewer bites per minute than did those in the NFC, $t(101.63) = 2.58$, $p = .011$, $d = 0.52$, 95% CI = [0.13, 0.91]. Participants in the NFC had 5.28 bites per minute (SD = 1.49), whereas those in FC had 4.55 bites per minute (SD = 1.40). In addition, participants in the FC had a higher success ratio than did those in the NFC, $t(98.87) = -4.13$, $p < .001$, $d = -0.89$, 95% CI = [-1.3, -0.49]. Participants in the FC consumed 66% of their bites outside the designated time interval, whereas those in the NFC consumed only 49% of their bites outside this interval.

However, these differences did not translate into a difference in the total amount of food consumed, $t(100.92) = -0.26$, $p = .797$, $d = -0.05$, 95% CI = [-0.43, 0.34]; participants in the FC consumed 435.77 grams of food (SD = 156.84) and participants in the NFC consumed 428.21 grams (SD = 141.38).

**Secondary outcomes**

A significant effect of condition on meal duration was found, $t(101.93) = -2.44$, $p = .016$, $d = -0.47$, 95% CI = [-0.86, -0.08]; participants in the FC consumed their meal in 9 minutes and 44 seconds, whereas those in the NFC consumed their meal in 8 minutes and 12 seconds. No differences between conditions were found in total fork servings, $t(99.55) = -0.03$, $p = .975$, $d = -0.01$, 95% CI = [-0.39, 0.38], or the average time interval between fork servings, $t(101.91) = -1.80$, $p = .074$, $d = -0.36$, 95% CI = [-0.75, 0.03]. Finally, participants in the FC did not report being more satiated after their meal than did those in the NFC, $t(96.4) = -0.24$, $p = .809$, $d = -0.05$, 95% CI = [-0.34, 0.44].

**Post-hoc analyses**

A significant effect of condition on grams of food consumed per minute was found, $t(101.54) = 2.1$, $p = .038$, $d = 0.43$, 95% CI = [0.04, 0.82]; participants in the FC consumed 48 grams per minute (SD = 21.94) whereas those in the NFC consumed 57.37 grams (SD = 23.46). No differences were found in average bite size between conditions, $t(101.27) = 0.54$, $p = .59$. In both conditions, participants consumed approximately 12 grams per bite.
After correcting for multiple testing for all p values reported above, only the effects of condition on total number of bites per minute (p = .017) and success ratio (p < .001) remain significant.

Discussion
This study examined the effect of real-time vibrotactile feedback delivered through the use of an augmented fork on eating rate, satiation, and food intake. It was expected that the participants who ate with a fork that provided vibrotactile feedback on their eating rate would take fewer bites per minute and take more bites outside the designated 10-second time interval than participants who did not receive feedback. It was further expected that that these changes in eating rate would lead to increased satiation and decreased ad-libitum food consumption. We found that participants who received feedback indeed had fewer bites per minute and consumed more bites outside the designated time interval of ten seconds. These changes, however, did not impact participants' satiation or food consumption.

The finding that real-time vibrotactile feedback delivered through an augmented fork reduces eating rate is consistent with literature on eating rate interventions that have utilized other forms of technology to modify eating behaviour (Ford et al., 2009; Ioakimidis et al., 2009; Zandian et al., 2009). The vibrotactile feedback delivered by the fork may have disrupted the automatic tendency to eat fast and may have served as a trigger to make alterations to one's eating rate (Hermsen, Frost, Renes, et al., 2016). Arguably, the feedback provided by the fork increases users' awareness of their eating rate. The real-time vibrotactile feedback enables users to compare their eating rate to their current goals (i.e. eating slower) and adapt their eating rate when their behaviour does not fit with their goals. Furthermore, it may also increase general self-awareness, which in turn increases one’s abilities to inhibit undesired behaviours (Alberts et al., 2011). Finally, it is known that among competing health-related behaviours, those supported by feedback are given priority over those without feedback (Northcraft et al., 2011). It is therefore conceivable that receiving vibrotactile feedback when eating too fast increased participants’ motivation to change their eating behaviour. The present findings demonstrate that real-time feedback delivered through digital technology may be an effective strategy to disrupt eating behaviour; even a very simple, non-intrusive type of feedback in the form of a simple vibration can function as a trigger for behaviour change and stimulate people to alter their eating rate.

Our results, however, failed to support the experimental hypothesis that a reduction in eating rate would lead to increased satiation and decreased ad-libitum food consumption. Although it has been shown that slower eating rate is
associated with lower energy intake, regardless of the type of manipulation used to change the eating rate such as type of instructions (Robinson, Almiron-Roig, et al., 2014), the context of the present study might explain why changes in eating rate did not translate into changes in satiation or energy intake. Firstly, although we could derive specific within-meal behaviours from the data gathered by the fork that are known to influence energy intake and/or satiation, such as bite speed and bite size (Andrade, Greene, & Melanson, 2008; Zijlstra et al., 2009), the fork was not specifically developed to modify other within-meal behaviours than the number of bites per minute. The fact that the fork did to specifically modify behaviours that have been shown to lower energy intake, such as oral processing time and number of chews per unit of food (Bolhuis et al., 2013a; Higgs & Jones, 2013; Weijzen et al., 2009), might explain the missing link between eating rate and reduced food intake in this study. Secondly, because it has been shown that there is a linear relationship between the size of experimental manipulation to eating rate (i.e. how much eating rate has been reduced by) and energy intake (Robinson, Almiron-Roig, et al., 2014), a further explanation as to why the reduction in eating rate observed in the present study did not reduce food intake is because the effect of decrease to eating rate was not large enough in size to impact food consumption. Thirdly, it is possible that because participants were asked to self-serve their meal size, participants cleared their plate out of habit rather than adjusting their intake based on eating rate or feeling of fullness. Therefore, it is possible that the initial effect of selected portion size may have overruled the effect of reducing eating rate (Brunstrom, 2011). Fourthly, it may be that specific characteristics of our test population have influenced our results. Our results demonstrated, for instance, that participants were not particularly motivated to change their eating rate in the near future. Feedback efficacy has been shown to be influenced by a high initial engagement with the target goal, i.e. reduction in eating rate, or strong motivation, i.e. to eat slower (Bandura, 1994). Although participants were found to eat slower in a response to the vibrotactile feedback, subsequently they may have not been motivated to eat less. To further understand the link between real-time vibrotactile feedback, eating rate and food intake, future research might examine whether and how initial motivation to change one’s eating rate or motivation to reduce food intake is affected by vibrotactile feedback. Finally, it has been argued that people may need to learn to associate the link between a slower eating rate, their satiety levels and energy intake (Brunstrom, 2011; Yeomans, Weinberg, & James, 2005). Although previous research has demonstrated the effects of a decelerated eating rate on food intake during a single meal (see Robinson, Almiron-Roig, et al., 2014), it is possible that receiving feedback would become effective across multiple meals. To test this
assumption, future studies may provide users with consistent feedback over a few meals and measure satiation and food intake over time.

A few limitations of the current study warrant discussion. Although the augmented fork seems a promising instrument to modify eating rate, more research is clearly warranted. The present study examined the effect of real-time vibrotactile feedback in a single sitting in a laboratory setting; therefore, the efficacy of the t0sFork in real-life settings is yet to be ascertained. To do so, we would encourage replication studies in ecologically-valid settings. It will be important for these studies to be adequately powered. Finally, because of the small variance in participants’ BMI, the current study could not test potential differences among normal-weight and overweight individuals in the extent to which their eating rate is affected by the vibrotactile feedback. Such an analysis would be a useful elaboration of the current research, given that differences in eating rate have been found between normal and overweight individuals (e.g. Ohkuma et al., 2015).

Taken together, the present study indicates that real-time vibrotactile feedback delivered through an augmented fork can reduce eating rate. Vibrotactile feedback led participants to eat fewer bites per minute and more bites outside the designated time interval of ten seconds. This indicates that vibrotactile feedback may be a viable tool to reduce eating rate. The changes in eating rate, however, did not translate into changes in satiation or energy intake. Future studies should examine the utility of the fork in real world settings, whether sustained use of the fork may result in decreased energy intake, and the utility of the fork with different test populations.
The effect of real-time vibrotactile feedback on eating rate, satiation, and energy intake
Chapter 6: Effects of a technology-based intervention to decelerate eating rate on eating rate and weight: a randomised controlled trial

Abstract

Eating rate is proposed to be a basic determinant of appetite regulation: people who eat more slowly feel sated earlier and eat less. As a result, fast eating rate may contribute to overeating and weight gain. Technology-based feedback offers a new way to decrease eating rate in naturalistic eating contexts. To assess the effects of a technology-based feedback intervention on participants’ eating rate and body weight over a 15-week period, we conducted a three-armed parallel group randomised controlled trial. A total of 141 participants with overweight or obesity (age = 49.15 ± 12.25; BMI (kg/m²) = 31.5 ± 4.48) were randomised to either one of two intervention groups (VFC, VFC+) or a control group (NFC). In VFC, participants received direct vibrotactile feedback from an augmented fork when eating too fast during a four-week training period. In VFC+, participants received the same vibrotactile feedback, but also had access to an online web portal with retrospective visual feedback on eating rate. In NFC, participants ate with the augmented fork without any form of feedback. Eating rate (i.e., success ratio (the percentage of bites with a sufficiently long pause between them) and bite rate) and body weight were measured at baseline (T1), directly after the 4-week training period (T2) and at 8-week follow-up (T3). Multilevel analysis showed that participants in both intervention groups had a significantly higher success ratio than those in the control group at T2. This effect persisted at T3. Bite rate only changed significantly at T2 for those in VFC. Participants in both intervention groups lost significantly more weight than those in the control group at T2 with no rebound at T3. This study showed that the use of an augmented fork to decrease eating rate may be an effective tool to reduce eating rate and promote weight loss.
Effects of a technology-based intervention to decelerate eating rate: a randomised controlled trial

Introduction

In recent decades, the prevalence of excessive body weight has increased rapidly in North-American and European countries, including The Netherlands (Hamann, 2017). In 2016, almost one of two Dutch adults were considered overweight (Volksgezondheidenzorg.info, 2018). Although a variety of factors are associated with overweight, evidence shows that eating quickly is positively associated with excess body weight (see Robinson, Almiron-Roig, et al. for a review), whereas a lower eating rate is associated with earlier feelings of satiation and a lower energy intake (Krop et al., 2018; Robinson, Almiron-Roig, et al., 2014). A promising method to prevent weight gain, therefore, may lie in encouraging those who eat quickly to slow down.

A potential barrier to changing eating rate may be its highly automatic habitual nature. In adults, eating rate is consistently found to be a personal characteristic, not depending on context (McCrickerd & Forde, 2017). In addition, research suggests that eating rate also has a heritable component (Llewellyn, Van Jaarsveld, Boniface, Carnell, & Wardle, 2008). Recent technological developments present new ways to measure and alter such highly automatic behaviors, by automatically sensing activity and providing feedback on undesired behaviors as they occur (Hermsen, Frost, Renes, et al., 2016).

A new and promising tool to alter eating rate is an augmented fork that contains sensors and actuators that provide real-time feedback on eating rate; when users of the fork eat too fast (i.e., taking more than one bite per 10 seconds), they feel a gentle vibration in the handle of the fork. This real-time vibrotactile feedback encourages people to slow down as they eat. The fork also provides retrospective visual feedback through a secure online dashboard, which is known to increase motivation to sustainably perform a desired behavior (Northcraft et al., 2011). Previous research suggests the fork is acceptable to users (Hermsen, Frost, Robinson, et al., 2016) and capable of reducing eating rate during a single meal in a laboratory context (Hermans et al., 2017).

The objective of the current work is to assess the longer-term effectiveness of the augmented fork in naturalistic eating contexts. Specifically, we aimed to examine the effects of both forms of feedback (i.e. vibrotactile and retrospective visual feedback) on participants’ eating rate and body weight after a four-week

This chapter is based on Hermsen, S., Mars, M., Higgs, S., Robinson, E., Frost, J.H., & Hermans, R.C.J. (2018). Effects of a technology-based intervention to decrease eating rate and body weight: a randomised controlled trial. A newer version of this manuscript and references to the final, peer-reviewed manuscript are available from the Take It Slow project OSF site: https://osf.io/753nf/

When referring to this chapter, please check the OSF site for the manuscript’s latest version.
training period with the augmented fork. To do so, we conducted a three-armed parallel group Randomised Controlled Trial (RCT) with two measures of eating rate (bite rate - average number of bites per minute - and success ratio - the percentage of bites with a sufficiently long pause between them) and body weight; these were measured at baseline (T1), directly after the 4-week training period (post-intervention, T2) and at a follow-up after eight weeks (T3). Based on the evidence regarding the effectiveness of feedback to disrupt habitual behavior (Hermsen, Frost, Renes, et al., 2016) and our previous work on the efficacy of the augmented fork to decrease eating rate (Hermans et al., 2017; Hermsen, Frost, Robinson, et al., 2016), we predicted that frequent use of the augmented fork would lead to a slower eating rate and this may translate to weight loss.

Methods

All of the participants were recruited by dieticians from their patient groups between November 2015 and April 2017. Using simple randomization procedures (computerized random numbers list generated with an online randomizer tool (Urbaniak & Pious, 2011); participants were assigned a number in order of enrollment in a single block; allocation was blinded to participants and their dieticians), we then assigned participants to 1 of three intervention groups: VFC (augmented fork measures eating rate and provides real-time vibrotactile feedback on behavior); VFC+ (augmented fork measures eating rate and provides both real-time vibrotactile feedback and retrospective visual feedback on behavior); and NFC (augmented fork measures eating rate; no feedback provided). The study was conducted between January 2016 and September 2017. The study was approved by the Ethics Committee of the Faculty of Social Sciences at Radboud University, Nijmegen, The Netherlands and was in full accordance with the Helsinki Declaration of 1975 as revised in 2013. All procedures involved were preregistered in the Netherlands Trial Register with number NTR5566. All participants provided their full written consent and received a gift voucher for €75 as compensation for participation.

Study sample

Dieticians from 30 practices recruited participants from their practices who met the inclusion criteria for the current study: 1) participants were at least 18 years old, 2) participants were self-reported fast eaters (see Table 1) and 3) participants had a BMI score ≥ 25 kg/m². Gastric bypass patients and were excluded from the study.
Our total sample consisted of 163 participants. An overview of the enrollment process and the allocation to intervention conditions is available in Figure 6.1.

**Intervention**

After eligibility assessment, participants received the augmented fork, along with an instruction manual for the download and installation of the software needed for synchronization, a unique login code for the software, and instructions for fork use and maintenance. Participants also received a leaflet with a briefing about how the fork measures eating rate, and the importance of mindful eating. Participants were then invited by email to complete an online baseline survey, hosted on the Qualtrics platform, in which they provided information on their gender, age, health condition, motivation to change eating rate, perceived eating rate, perceived detriments of eating rate, perceived satiety, awareness of meal, and awareness of eating rate. Furthermore, participants were weighed and their height was measured by their dieticians, using standardized equipment and procedures at the dietician's practice.

We then assessed participants' eating rate during a baseline measurement (T1). All participants ate as many meals as possible with the fork during a period of five consecutive days. In this period, participants did not receive any form of feedback on their eating rate. Participants received no instructions on a minimum or maximum number of meals, or limitations on where they could use the fork, to minimize interference with their natural eating habits.

After establishing this baseline measure, we randomly assigned participants to one of three conditions:

1. **Augmented fork with real-time vibrotactile feedback (VFC condition)**
2. **Augmented fork with both real-time vibrotactile feedback and retrospective visual feedback (VFC+ condition)**
3. **Augmented fork without any kind of feedback, neither vibrotactile nor visual (NFC condition)**

In all conditions, participants then entered the intervention phase, consisting of a four-week training period in which all participants were asked to eat as many meals as possible with the fork. All participants once more received instructions on the importance of eating slowly and the beneficial effects of mindful eating. Furthermore, participants in VFC+ received instructions about the way the fork provided vibrotactile feedback on eating rate. Additionally, participants in VFC+ received an invitation to visit an online website that provided visual retrospective feedback on their eating rate. This online dashboard was blocked for participants in NFC and VFC. During the four-week training, all participants ate as many meals as possible with the rosFork.

After the training period, all participants entered a post-training measurement (T2), in which they used the fork without any form of feedback for five consecutive
days, eating as many meals as possible with the fork to establish their post-
training eating rate. Moreover, they were re-weighed by their dieticians.

Participants then entered a period of eight weeks in which they could not use
the fork. After this period, in a follow-up measurement (T3), participants in all
three conditions were once again re-weighed by their dieticians and used the fork
without any form of feedback for five consecutive days, to test for sustainable
changes in eating rate. Figure 6.1 (CONSORT flowchart) provides an overview of
the procedure and experimental design of the study.

Stimulus materials
Feedback on eating rate was provided by the 10sFork, developed and marketed by
SlowControl (Paris, France). This augmented fork has the appearance of a regular
fork but contains sensors and actuators that provide real-time feedback on eating
rate (see Chapter 4 of this thesis for a detailed description of the fork). The fork
delivers feedback at a pre-set interval between bites; in this study, the interval was
set to the factory default of 10 seconds. If users take a bite too quickly (i.e. before
the end of the 10-second interval), they feel a gentle vibration in the handle of the
fork and see a red indicator light. The fork stores each ‘bite’ with a unique ID and
timestamp, which enables determining meal duration through the exact time at
which the meal is started and ended. Furthermore, it counts the total number of
bites per meal and per minute, and the average interval between bites. Finally, it
measures the ratio of ‘correct’ bites versus bites within the 10s timeframe. All data
is stored on the fork and can be synchronized with an online platform. In addition
to the vibrotactile feedback, (only the) participants in VFC+ had access to a secure
online platform, where they could review retrospective visual feedback on (trends
in) meal duration, number of bites, and over-speed ratio of their past meals.

Measures

Primary outcome measures
Bite rate and success ratio. For every participant, the 10sFork was set up to automati-
cally record each bite of each individual meal. For each unique meal, participants’
bite rate (i.e., the average number of bites per minute ((total number of bites
divided by meal duration in seconds) multiplied by 60) and success ratio (i.e.,
number of bites outside 10-second time interval divided by total bites) were calcu-
lated automatically by a script on the SlowConnect server.

Body weight. Participants’ weight and height were measured by their dieti-
cians, following standard procedures, at three moments: at baseline (T1), directly
after the 4-week training period (T2) and at the follow up after eight weeks (T3).
Enrollment

Recruited by dieticians
\(n = 163\)

Assessed for eligibility
\(n = 163\)

Excluded: \(n = 0\)
Not meeting inclusion criteria: \(n = 0\)
Declined to participate after receiving 10sFork, but before start of trial: \(n = 12\)
Failed to participate (technological issues, etc.): \(n = 10\)

T1: Baseline measurement

T1: Pre-measurement of eating rate and BMI, Q1
\(n = 141\)

Allocation

Randomized: \(n = 141\)

Allocated to intervention:
Vibrotactile Feedback Condition (VFC), \(n = 51\)
Received allocated intervention: \(n = 45\)
Did not receive intervention: \(n = 6\)

Allocated to intervention:
Vibrotactile + Visual Feedback Condition (VFC+), \(n = 44\)
Received allocated intervention: \(n = 39\)
Did not receive intervention: \(n = 5\)

Allocated to intervention:
Control / No Feedback Condition (NFC), \(n = 46\)
Received allocated intervention: \(n = 36\)
Did not receive intervention: \(n = 10\)

T2: Post-measurement after 4 weeks training

T2: Post-measurement of eating rate, success ratio, and BMI, Q2
\(n = 120\) (VFC 45, VFC+ 39, NFC 36)

T3: Follow-up measurement after eight weeks

T3: Post-measurement of eating rate, success ratio, and BMI
\(n = 100\) (VFC 33, VFC+ 35, NFC 32)

Unavailable for follow-up:
VFC \(n = 12\), VFC+ \(n = 4\), NFC \(n = 4\)

Analysis

Analysed
\(n = 120\) (of which 100 with complete data), excluded from analysis: \(n = 0\)

Figure 6.1. CONSORT flowchart of enrollment, allocation, and experimental design
Participants’ BMI was calculated as weight in kilograms divided by height in meters squared.

Secondary outcome measures

Meal duration. Meal duration was calculated as the time in minutes between the first and last bite of every unique meal.

Pause duration. The average time interval between fork servings (i.e. pause duration), was determined by calculating the average interval between bites for each unique meal.

Total bites. The total number of fork servings for each unique meal was calculated by a script on the SlowConnect server.

Randomisation checks and potential confounders

We determined participants’ age†, gender†, health condition†† (dietary restraint, diabetes I and II, stomach complaints (heartburn, regurgitation, bloating, obstipation, flatulence), discomfort during eating), eating behaviour††† (defined as aspects of the Dutch Eating Behaviour Questionnaire (DEBQ): restrained eating, emotional eating, and external eating, Van Strien et al., 1986), motivation††††, perceived self-efficacy††††, perceived efficacy of the fork††††, perceived eating rate at baseline††††, perceived detriments of eating rate†††† and awareness of eating†††† through questions taken from the online baseline survey².

Statistical analyses

We inspected the sample distributions and distributions of the mean of our primary and secondary outcome variables and their potential confounders (those determinants named above under randomization checks, plus BMI at baseline), testing skewness, kurtosis, and performing Hartigan’s dip test (Hartigan & Hartigan, 1985), Shapiro Wilk-test for normality (Wilk & Shapiro, 1968), Anderson-Darling-test for goodness of fit (Anderson & Darling, 1954), and Kolmogorov-Smirnov test for equality of distributions (Smirnov, 1948). We then performed randomization checks for group equivalence, calculating Cramér’s V for categorical variables, and Omega squared for continuous variables. Finally, we checked whether BMI and secondary outcome measures – meal duration, pause duration, and total bites – were associated at baseline with bite rate and success ratio.

To assess the short- and medium-term effectiveness of the augmented fork on our primary (bite rate, success ratio, and BMI) and secondary (meal duration, pause duration, total bites) outcome measures over time, we performed a multilevel analysis. Multilevel analysis is best suited to deal with clustering of data and

² †: open questions, ††: digital choice (yes / no), †††: seven-point scale 1–7, ††††: ten-point scale 1–10
Effects of a technology-based intervention to decelerate eating rate: a randomised controlled trial

different baselines, and incomplete cases without the need for pairwise or listwise deletion of cases, or potentially problematic imputation of missing values (Baayen, Davidson, & Bates, 2008). For each outcome measure, we tested a model with phase (T1, T2 and T3) and condition (NFC, VFC, VFC+) as fixed effects, and an intercept for subject as random effect. We report mean estimates, standard deviations, significance level, effect sizes (Cohen’s $d$) and 95% confidence interval for the effect sizes for each outcome measure. All analyses were done in R version 3.3.2 for MacOS X with RStudio version 1.0.136 for MacOS X (R Core Team, 2017), using the lme4 package (Bates, Machler, Bolker, & Walker, 2015) version 1.1-15 for the multi-level analyses, with additional t-tests done with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) version 2.0-36 to determine Satterthwaite approximations to degrees of freedom.

Statistical power calculations to determine the sample size needed for multi-level modelling tend to be more complex than those needed for single-level designs, because of the statistical dependency of clustered data. Rules of thumb for sample size per cell on the lowest level range from 30–50 participants per cell (Kreft & De Leeuw, 1998; Maas & Hox, 2005). With three cells in a single level (condition), and three repeated observations per participant, we needed to include 90–150 participants in our trials to obtain 80% power to detect a medium effect size (as found in Hermans et al., 2017), with a significance level of 0.05, assuming the ratio of the variability of the level 1 coefficient to the variability of the level 1 residual is 1.

A complete overview of all analyses, R code use, and results are available in the Analysis Report through the Radboud Repository Open Science platform.

Results

Of the total number of 163 participants, 141 were available for allocation to one of three conditions: intervention conditions VFC and VFC+, and control condition NFC (see Figure 6.1). Table 6.1 shows the baseline characteristics of the subjects in the intervention and control conditions. All tests for sample distributions, group equivalence, and associations at baseline revealed no irregularities. Of the 141 participants allocated to the three conditions, 120 (85.1%) were retained in the trial at T2 (the post-measurement period directly after the training, VFC 45, VFC+ 39, NFC 36); at T3, 100 participants (70.9%) still took part (VFC 33, VFC+ 35, NFC 32).

During the training period, participants ate an average of 27.8 meals ($SD = 16.3$) with the 10sFork in 4 weeks, an average of one meal per day. Patterns for the amount of meals eaten with the fork during training were similar for each condition (VFC $m = 25.6 \pm 14.7$; VFC+ $m = 28.9 \pm 18.0$; NFC $m = 29.7 \pm 16.5$, $\omega^2=.008$).
Chapter 6

Of the 39 participants in the VFC+ condition, 21 used the online dashboard for additional retrospective visual feedback on their eating rate: four participants used the dashboard more than once per week, 11 used it once per week at most, 3 used it a couple of times in the training period, and 3 more used it only once in the 4-week training period.

### Table 6.1
Baseline characteristics of study participants across three conditions

<table>
<thead>
<tr>
<th>Variable</th>
<th>NFC (n = 51)</th>
<th>VFC (n = 44)</th>
<th>VFC+ (n = 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y (SD)</td>
<td>50.6 (11.0)</td>
<td>49.6 (11.0)</td>
<td>47.40 (14.4)</td>
</tr>
<tr>
<td>Female, n (% of respondents)</td>
<td>21 (58% of 36)</td>
<td>26 (59% of 44)</td>
<td>22 (65% of 34)</td>
</tr>
<tr>
<td>BMI, kg/m² (SD)</td>
<td>31.7 (5.2)</td>
<td>31.2 (4.4)</td>
<td>31.6 (3.9)</td>
</tr>
<tr>
<td>Perceived eating rate (SD)</td>
<td>7.7 (1.3)</td>
<td>7.6 (1.1)</td>
<td>7.7 (1.3)</td>
</tr>
<tr>
<td>Perceived eating discomfort (%)</td>
<td>None: 35 (97%)</td>
<td>None: 43 (98%)</td>
<td>None: 46 (100%)</td>
</tr>
<tr>
<td>Stomach complaints (%)</td>
<td>18 (47%)</td>
<td>26 (55%)</td>
<td>20 (53%)</td>
</tr>
<tr>
<td>Has diabetes I (%)</td>
<td>3 (9%), 2 (6%)</td>
<td>2 (5%), 9 (21%)</td>
<td>5 (14%), 0 (0%)</td>
</tr>
<tr>
<td>Is on a diet (%) of Q1 respondents</td>
<td>10 (29%) of 36</td>
<td>15 (34%) of 44</td>
<td>15 (43%) of 34</td>
</tr>
</tbody>
</table>

1 – NFC, no feedback condition; VFC, vibrotactile feedback condition, VFC+, vibrotactile + visual feedback condition

2 – All data are for those participants allocated to the intervention (NFC n = 51; VFC, n = 44, VFC+, n = 46) AND filled out the premeasurement questionnaire (NFC n = 36; VFC, n = 44, VFC+, n = 34)

3 – On a scale from 1–10

4 – Dichotomous scale: Yes / No

### Primary outcome measures

#### Bite rate

An interaction effect between phase (T1, T2 or T3) and condition (NFC, VFC or VFC+) on average bite rate per minute was found when comparing the NFC group with the VFC group at T2 (Estimate = -1.08 bites/minute ± 0.49, t (329.0) = -2.217, p < 0.05, see Table 6.3 for a full overview of multilevel analyses of primary outcome measures). No significant interactions between phase and condition occurred when comparing the NFC group with the VFC group at T3, nor when comparing the NFC group to the VFC+ group at T2 and T3 (see Table 6.3). A main effect for phase at T2 (Estimate = -0.79 bites /minute ± 0.36, t (328.9) = -2.189, p = .03) was qualified by the interaction between phase and condition at T2 for the VFC group. The effect size (Cohen’s d) for the comparison between NFC and VFC at T2 was 0.79, indicating a moderate to large effect (see Table 6.4 for an overview of effect sizes for experimental conditions compared to control).

On average, participants in the VFC condition took 6.20 ± 2.60 bites per minute at T1, slowing down to 4.35 ± 1.65 bites per minute at T2 and 4.85 ± bites per minute at T3 (see Table 6.2 for means, standard deviations, and 95% confidence intervals for all primary outcome measures, per condition). Participants in the NFC
Table 6.2
Means and standard deviations of primary outcome measures, per condition

<table>
<thead>
<tr>
<th></th>
<th>NFC</th>
<th>VFC</th>
<th>VFC+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bite Rate: Baseline</strong></td>
<td>5.72 ± 6.44</td>
<td>5.55 ± 6.20</td>
<td>5.08 ± 5.81</td>
</tr>
<tr>
<td>T2</td>
<td>4.94 ± 5.65</td>
<td>3.67 ± 4.35</td>
<td>3.45 ± 4.18</td>
</tr>
<tr>
<td>T3</td>
<td>5.44 ± 6.14</td>
<td>4.30 ± 4.85</td>
<td>3.99 ± 4.63</td>
</tr>
<tr>
<td><strong>Success Ratio: Baseline</strong></td>
<td>38.2 ± 45.1</td>
<td>35.7 ± 44.7</td>
<td>39.6 ± 46.3</td>
</tr>
<tr>
<td>T2</td>
<td>37.9 ± 44.5</td>
<td>58.5 ± 64.3</td>
<td>60.2 ± 66.9</td>
</tr>
<tr>
<td>T3</td>
<td>34.8 ± 42.7</td>
<td>46.3 ± 54.9</td>
<td>49.9 ± 58.3</td>
</tr>
<tr>
<td><strong>BMI: Baseline</strong></td>
<td>30.4 ± 31.9</td>
<td>29.8 ± 31.2</td>
<td>30.1 ± 31.6</td>
</tr>
<tr>
<td>T2</td>
<td>30.6 ± 32.1</td>
<td>29.3 ± 30.6</td>
<td>29.6 ± 30.3</td>
</tr>
<tr>
<td>T3</td>
<td>30.0 ± 31.5</td>
<td>29.0 ± 29.8</td>
<td>29.4 ± 30.7</td>
</tr>
</tbody>
</table>

1 – T2 = post-measurement after training period; T3 = follow-up measurement after eight weeks; NFC, no feedback condition; VFC, vibrotactile feedback condition; VFC+, vibrotactile + visual feedback condition

2 – Bite rate: lower 95% confidence interval < average number of bites per minute ± SD < upper 95% confidence interval; Success ratio: lower 95% confidence interval < percentage of bites with at least second pause ± SD < upper 95% confidence interval; BMI: lower 95% confidence interval < kg/m² ± SD < upper 95% confidence interval.

Table 6.3
Multilevel analysis of intervention effect on eating rate, success ratio, and BMI over time

<table>
<thead>
<tr>
<th></th>
<th>Success Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bite Rate</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>6.44 ± 3.36</td>
</tr>
<tr>
<td>T2</td>
<td>-0.79 ± 3.36</td>
</tr>
<tr>
<td>T3</td>
<td>-0.22 ± 3.36</td>
</tr>
<tr>
<td>VFC</td>
<td>-0.24 ± 3.36</td>
</tr>
<tr>
<td>VFC+</td>
<td>-0.63 ± 3.36</td>
</tr>
<tr>
<td>T2*VFC</td>
<td>-1.08 ± 3.36</td>
</tr>
<tr>
<td>T3*VFC</td>
<td>-0.95 ± 3.36</td>
</tr>
<tr>
<td>T2*VFC+</td>
<td>-0.84 ± 3.36</td>
</tr>
<tr>
<td>T3*VFC+</td>
<td>-0.85 ± 3.36</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>31.87 ± 3.36</td>
</tr>
<tr>
<td>T2</td>
<td>0.22 ± 3.36</td>
</tr>
<tr>
<td>T3</td>
<td>-0.37 ± 3.36</td>
</tr>
<tr>
<td>VFC</td>
<td>-0.72 ± 3.36</td>
</tr>
<tr>
<td>VFC+</td>
<td>-0.30 ± 3.36</td>
</tr>
<tr>
<td>T2*VFC</td>
<td>-0.73 ± 3.36</td>
</tr>
<tr>
<td>T3*VFC</td>
<td>-0.44 ± 3.36</td>
</tr>
<tr>
<td>T2*VFC+</td>
<td>-0.64 ± 3.36</td>
</tr>
<tr>
<td>T3*VFC+</td>
<td>-0.26 ± 3.36</td>
</tr>
</tbody>
</table>

1 – T2 = post-measurement after training period; T3 = follow-up measurement after eight weeks; NFC, no feedback condition; VFC, vibrotactile feedback condition; VFC+, vibrotactile + visual feedback condition

2 – Values are mean estimates ± SEs, t-values with Satterthwaite approximation of degrees of freedom. Significance * p < .05, ** p < .01, *** p < .001.
condition took (on average) 6.44 ± 2.50 bites per minute at T1, 5.65 bites / minute ± 2.68 at T2, and 6.14 ± 2.60 bites / minute at T3; participants in VFC+ took (on average) 5.81 bites / minute ± 2.60 at T1, 4.18 bites / minute ± 2.60 at T2, and 4.63 ± 2.14 bites / minute at T3.

Success ratio

We found an interaction effect between phase and condition at T2 when comparing the NFC group with the VFC group (Estimate = 23.1 % ± 4.5, t (332.2) = 5.132, p < .001), and when comparing the NFC group with the VFC+ group (Estimate = 20.9 % ± 4.7, t (331.7) = 4.432, p < .001. Furthermore, we found an interaction between phase and condition at T3 when comparing the NFC group with the VFC group (Estimate = 13.4 % ± 5.1, t (337.2) = 2.646, p < .01). Effect sizes (Cohen’s d) for the comparison between NFC and VFC at T2 and T3 were 0.86 and 0.52 respectively, and effect sizes for the comparison between NFC and VFC+ at T2 and T3 were 1.11 and 0.71 respectively (see Table 6.4 for an overview of effect sizes for experimental conditions compared to control).

Participants in the NFC condition had an average success ratio of 45.1% ± 21.1 at T1, 44.5% ± 20.1 at T2, and 42.7% ± 21.4 at T3. Participants in the VFC condition had an average success ratio of 41.7% ± 20.5 at T1, 64.3% ± 23.0 at T2, and 54.9% ± 23.5 at T3. Participants in the VFC+ condition had an average success ratio of 46.3% ± 18.6 at T1, 66.9% ± 22.1 at T2, and 58.3% ± 22.0 at T3 (see Table 6.2 for a full overview of means, standard deviations, and 95% confidence intervals).

Body weight (BMI)

We found an interaction effect between phase and condition when comparing the NFC group to the VFC group at T2 (Estimate = -0.73 kg/m² ± 0.23, t (212.99) = -3.169, p < .01) and when comparing the NFC group to the VFC+ group at T2 (Estimate = -0.64 kg/m² ± 0.24, t (212.9500) = -2.654, p < .01). No significant interaction effects between phase and condition occurred at T3, but we did find a main effect of phase on BMI at T3 (Estimate = -0.37 kg/m² ± 0.17, t (212.95) = -2.148, p < 0.05). All participants regardless of condition appeared to have lost some weight between T2 and T3 (but see below): average BMI of all participants went from 31.18 kg/m² ± 4.73 at T2 to 30.62 kg/m² ± 4.24 at T3. This corresponded with an average weight loss of 1.19 kg (NFC: 0.60 kg, VFC: 1.75 kg, VFC+: 1.05kg).

Effect sizes (Cohen’s d) for the comparison at T2 between NFC and VFC and between NFC and VFC+ were 0.33 and 0.20 respectively, indicating a small to moderate effect (see Table 6.4). Participants in the NFC condition had an average BMI of 31.9 kg / m² ± 5.2 at T1, 32.1 kg / m² ± 5.7 at T2, and 31.5 kg / m² ± 5.0 at T3. Participants in the VFC group had an average BMI of 31.2 kg / m² ± 4.4 at T1, 30.6 kg
Effects of a technology-based intervention to decelerate eating rate: a randomised controlled trial

/ m² ± 4.5 at T2, and 29.8 kg / m² at T3; i.e. they reduced their BMI by an average of 0.56 kg/m² at T2, with a further reduction of 0.84 kg/m² at T3, which corresponded with an average total weight loss of 1.23 kg at T2 and 2.98 kg at T3. Participants in VFC+ had an average BMI of 31.6 kg/m² ± 3.9 at T1, 31.0 kg / m² ± 3.8 at T2, and 30.7 kg / m² ± 4.0 at T3; i.e. they reduced their BMI by an average of 0.62 points at T2, with a further reduction of 0.33 kg/m² at T3, which corresponded with an average total weight loss of 1.83 kg at T2 and 2.88 kg at T3 (see Table 6.2 for a full overview of means, standard deviations, and 95% confidence intervals).

Table 6.4
Effect sizes of comparisons between experimental conditions (VFC and VFC+) with the control condition (NFC) for each phase, for bite rate, success ratio, and BMI

<table>
<thead>
<tr>
<th></th>
<th>Bite Rate</th>
<th>Success Ratio</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: NFC–VFC</td>
<td>-0.33 &lt; 0.09 &lt; 0.52</td>
<td>-0.26 &lt; 0.17 &lt; 0.59</td>
<td>-0.28 &lt; 0.16 &lt; 0.60</td>
</tr>
<tr>
<td>T1: NFC–VFC+</td>
<td>-0.20 &lt; 0.71 &lt; 0.25</td>
<td>-0.51 &lt; 0.06 &lt; 0.39</td>
<td>-0.38 &lt; 0.08 &lt; 0.54</td>
</tr>
<tr>
<td>T1: VFC–VFC+</td>
<td>-0.25 &lt; 0.58 &lt; 0.15</td>
<td>-0.65 &lt; 0.22 &lt; 0.20</td>
<td>-0.54 &lt; 0.10 &lt; 0.35</td>
</tr>
<tr>
<td>T2: NFC–VFC</td>
<td>0.34 &lt; 0.79 &lt; 1.23</td>
<td>-1.30 &lt; -0.86 &lt; 0.41</td>
<td>-0.11 &lt; 0.33 &lt; 0.78</td>
</tr>
<tr>
<td>T2: NFC–VFC+</td>
<td>0.09 &lt; 1.01 &lt; 0.55</td>
<td>-1.59 &lt; -1.11 &lt; 0.63</td>
<td>-0.27 &lt; 0.20 &lt; 0.67</td>
</tr>
<tr>
<td>T2: VFC–VFC+</td>
<td>-0.33 &lt; 0.10 &lt; 0.53</td>
<td>-0.54 &lt; -0.11 &lt; 0.32</td>
<td>-0.53 &lt; -0.08 &lt; 0.37</td>
</tr>
<tr>
<td>T3: NFC–VFC</td>
<td>0.05 &lt; 0.56 &lt; 1.07</td>
<td>-1.03 &lt; -0.52 &lt; 0.01</td>
<td>0.03 &lt; 0.50 &lt; 0.96</td>
</tr>
<tr>
<td>T3: NFC–VFC+</td>
<td>0.20 &lt; 0.71 &lt; 1.22</td>
<td>-1.22 &lt; -0.71 &lt; -0.20</td>
<td>-0.32 &lt; 0.16 &lt; 0.65</td>
</tr>
<tr>
<td>T3: VFC–VFC+</td>
<td>-0.38 &lt; 0.10 &lt; 0.58</td>
<td>-0.64 &lt; -0.16 &lt; 0.32</td>
<td>-0.74 &lt; -0.27 &lt; 0.19</td>
</tr>
</tbody>
</table>

1 – T2 = post-measurement after training period; T3 = follow-up measurement after eight weeks; NFC, no feedback condition; VFC, vibrotactile feedback condition, VFC+, vibrotactile + visual feedback condition
2– Values are lower 95% confidence interval < point estimate for Cohen’s d < upper 95% confidence interval

Secondary outcome measures

For meal duration, we found main effects of phase at T2 (Estimate = 95.0 seconds ± 43.7, t (333.3) = 2.177, p < .05) and T3 (Estimate = 97.3 seconds ± 48.1, t (342.2) = 2.023, p < .05). Furthermore, we found a main effect of condition for the VFC group (Estimate = 122.4 seconds ± 56.6, t (289.2) = 2.164, p < .05). All participants spent more time on their meals at T2 (690.5 seconds ± 360.5) and T3 (733.4 seconds ± 453.5) than at baseline (T1: 625.2 seconds ± 352.4). Moreover, participants in the VFC condition, on average, spent more time on their meals in all three measurement phases than participants in NFC and VFC+ (see Table 6.5).

We found an interaction effect of phase and condition on pause duration at T2; when compared with the NFC group, participants in both the VFC (Estimate = 3.1 seconds ± 1.6, t (333.1) = 1.985, p < 0.05) and VFC+ (Estimate = 4.3 seconds ± 1.7, t (332.7) = 2.635, p < 0.01) groups took longer breaks between bites after the training than participants in NFC. Participants in VFC, on average, prolonged their breaks between bites from 13.5 ± 6.1 seconds at T1 to 18.2 ± 8.6 seconds at T2. Participants in VFC+, showed similar results, with average breaks of 14.7 ± 6.9 seconds at T1 and
Chapter 6

Table 6.5
Means and standard deviations of secondary outcome measures, per condition

<table>
<thead>
<tr>
<th></th>
<th>NFC</th>
<th>VFC</th>
<th>VFC+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meal duration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>505 &lt; 589.9 ± 164.1 &lt; 670</td>
<td>635 &lt; 709.8 ± 325.2 &lt; 784</td>
<td>593 &lt; 676.1 ± 281.6 &lt; 759</td>
</tr>
<tr>
<td>T2</td>
<td>600 &lt; 682.4 ± 212.9 &lt; 764</td>
<td>645 &lt; 721.1 ± 271.1 &lt; 705</td>
<td>591 &lt; 673.8 ± 213.8 &lt; 757</td>
</tr>
<tr>
<td>T3</td>
<td>594 &lt; 659.7 ± 255.5 &lt; 775</td>
<td>653 &lt; 733.9 ± 224.4 &lt; 825</td>
<td>701 &lt; 792.7 ± 477.7 &lt; 873</td>
</tr>
<tr>
<td><strong>Pause duration</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>10.6 ± 13.4 ± 5.5 &lt; 15.4</td>
<td>11.1 ± 13.5 ± 6.1 &lt; 15.6</td>
<td>12.3 ± 14.7 ± 6.9 &lt; 17.1</td>
</tr>
<tr>
<td>T2</td>
<td>12.3 ± 14.7 ± 6.2 &lt; 17.0</td>
<td>16.1 ± 18.2 ± 8.6 &lt; 20.4</td>
<td>18.3 ± 20.7 ± 10.6 &lt; 23.3</td>
</tr>
<tr>
<td>T3</td>
<td>12.4 ± 14.9 ± 7.0 &lt; 17.5</td>
<td>14.3 ± 16.8 ± 7.3 &lt; 19.1</td>
<td>15.8 ± 18.9 ± 9.6 &lt; 20.7</td>
</tr>
<tr>
<td><strong>Total bites</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>47.7 ± 56.2 ± 22.7 &lt; 64.0</td>
<td>58.4 ± 65.8 ± 32.7 &lt; 73.2</td>
<td>51.8 ± 60.1 ± 26.7 &lt; 68.3</td>
</tr>
<tr>
<td>T2</td>
<td>51.1 ± 59.2 ± 28.8 &lt; 67.3</td>
<td>41.5 ± 49.3 ± 24.2 &lt; 56.3</td>
<td>34.0 ± 42.2 ± 20.0 &lt; 50.4</td>
</tr>
<tr>
<td>T3</td>
<td>55.8 ± 60.7 ± 25.3 &lt; 73.2</td>
<td>50.2 ± 55.9 ± 31.7 &lt; 66.3</td>
<td>45.2 ± 52.3 ± 22.2 &lt; 62.1</td>
</tr>
</tbody>
</table>

1 – Meal duration: lower 95% CI < seconds ± SD < higher 95% CI;
Pause duration: lower 95% CI < seconds ± SD < higher 95% CI;
Total bites: lower 95% CI < bites per meal ± SD < higher 95% CI

Table 6.6
Multilevel analysis of intervention effect on secondary outcome measures over time

<table>
<thead>
<tr>
<th>Meal duration</th>
<th>Pause duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>587.4 ± 42.0, t (291.8) = 13.091, p = .000 ***</td>
</tr>
<tr>
<td>T2</td>
<td>95.0 ± 43.7, t (333.3) = 2.177, p = .03 *</td>
</tr>
<tr>
<td>T3</td>
<td>97.3 ± 48.1, t (342.2) = 2.023, p = .04 *</td>
</tr>
<tr>
<td>VFC</td>
<td>122.4 ± 56.6, t (289.2) = 2.164, p = .03 *</td>
</tr>
<tr>
<td>VFC+</td>
<td>88.7 ± 59.5, t (288.9) = 1.491, p = .13</td>
</tr>
<tr>
<td>T2*VFC</td>
<td>-84.0 ± 59.2, t (333.3) = -1.439, p = .15</td>
</tr>
<tr>
<td>T3*VFC</td>
<td>-97.3 ± 64.1, t (342.5) = -1.956, p = .29</td>
</tr>
<tr>
<td>T3*VFC+</td>
<td>-17.3 ± 66.1, t (283.2) = 0.207, p = .84</td>
</tr>
<tr>
<td>Total bites</td>
<td>55.86 ± 4.13, t (215.6) = 13.530, p = .000 ***</td>
</tr>
<tr>
<td>T2</td>
<td>3.36 ± 3.54, t (333.1) = 0.950, p = .34</td>
</tr>
<tr>
<td>T3</td>
<td>8.60 ± 3.91, t (336.3) = 2.202, p = .03 *</td>
</tr>
<tr>
<td>VFC</td>
<td>9.94 ± 5.57, t (213.8) = 1.784, p = .08</td>
</tr>
<tr>
<td>VFC+</td>
<td>4.28 ± 5.86, t (213.6) = 0.730, p = .47</td>
</tr>
<tr>
<td>T2*VFC</td>
<td>-20.29 ± 4.78, t (331.2) = -4.243, p = .000 ***</td>
</tr>
<tr>
<td>T3*VFC</td>
<td>-16.17 ± 5.32, t (336.5) = -3.043, p = .003 **</td>
</tr>
<tr>
<td>T2*VFC+</td>
<td>-21.32 ± 5.01, t (330.8) = -4.252, p = .000 ***</td>
</tr>
<tr>
<td>T3*VFC+</td>
<td>-15.09 ± 5.37, t (334.9) = -2.810, p = .005 **</td>
</tr>
</tbody>
</table>

1 – T2 = post-measurement after training period; T3 = follow-up measurement after eight weeks; NFC, no feedback condition; VFC, vibrotactile feedback condition, VFC+, vibrotactile + visual feedback condition
Values are mean estimates ± SEs, t-values with Satterthwaite approximation of degrees of freedom.
Significance * p < .05; ** p < .01; *** p < .001.

2 – Values are mean estimates ± SEs, t-values with Satterthwaite approximation of degrees of freedom.
Significance * p < .05; ** p < .01; *** p < .001.
20.7 ± 10.6 seconds at T2. Participants in NFC did not show longer breaks between bites: 13.1 ± 5.5 seconds at T1, 14.7 ± 6.2 seconds at T2.

An interaction of phase and condition on total bites occurred at both T2 and T3. When compared to participants in NC, participants in VFC took less bites per meal at both T2 (Estimate = -20.3 bites per meal ± 4.8, t (331.2) = -4.243, p < .001) and T3 (Estimate = -16.2 bites per meal ± 5.3, t (336.5) = -3.043, p < .01). Participants in VFC+ showed similar results when compared to NFC. Both at T2 (Estimate = -21.3 bites per meal ± 5.0, t (330.8) = -4.252, p < .001) and T3 (Estimate = -15.1 bites per meal ± 5.4, t (334.5) = -2.810, p < .001) they took significantly fewer bites per meal.

All means, standard deviations, and 95% confidence intervals for the secondary outcome measures at T1, T2, and T3 are reported in Table 6.5. All multilevel analyses for the secondary outcome measures are reported in Table 6.6.

**Discussion**

The present study assessed the effects of a technology-based feedback intervention on eating rate and body weight in naturalistic eating contexts. Specifically, we examined the effects of vibrotactile and retrospective visual feedback on participants’ eating rate and body weight over a 15-week period. To do so, we conducted a three-armed parallel group RCT with two measures of eating rate (i.e., bite rate and success ratio) and body weight, measured at baseline, directly after a 4-week training period (post-intervention, T2) and eight weeks later (follow-up, T3).

Our findings show that technology-based feedback helps to sustainably slow down eating rate. The effect of this feedback mostly affects success ratio, i.e. the ‘spacing’ of bites during a meal, and much less (if at all) the total average of bites per minute during the entire meal (bite rate). This suggests that the feedback delivered by the fork teaches people to slow down their ‘fast’ bites. Our findings on pause duration confirm this result: people in the intervention conditions had significantly longer pauses between bites. This is only partially in line with our hypothesis that people who received concurrent vibrotactile feedback on eating rate would succeed in eating slower, with both a lower average number of bites per minute (bite rate), and more time between bites (success ratio). Our results confirm and further augment findings from a previous (lab) study with the 10sFork (Hermans et al., 2017), in which feedback from the fork had a significant effect on both bite rate and success ratio in a single setting, but with the latter effect much more pronounced. The finding that the effect on success ratio remains significant at an eight-week follow-up measurement shows that feedback from digital technology has the potential to lastingly change eating rate.
Chapter 6

All participants, regardless of condition, had longer meals. This may very well be an effect of the demand characteristics of eating with the 10sFork, which in itself encourages more mindful eating. This is in line with previous findings (Hermsen, Frost, Robinson, et al., 2016) where people reported that eating with the fork made them more aware of their eating behavior. However, only in the intervention groups did eating with the fork lead to a higher success ratio, longer pauses between bites, and less bites per meal.

In the intervention groups that received vibrotactile feedback on eating rate, we found an effect of the feedback on participants’ weight, which persisted after the eight-week period. Directly after the training, participants in the intervention (VFC and VFC+) groups managed to lose weight, whereas participants in the control group did not. Moreover, at the follow-up measurement after eight weeks, participants in the intervention groups had not regained their lost body weight. This finding is in line with other feedback-based interventions to reduce eating rate, which, in clinical settings, also found durable weight loss among adults (Bergh et al., 2008; Bergh et al., 2013) and children (Ford et al., 2009).

Our finding that feedback on eating rate leads to weight loss may be explained by a number of potential underlying mechanisms. The fact that participants in the intervention groups took less bites per meal may have meant that they also ate less, and therefore took in less calories. Furthermore, eating rate may have influenced intake through such mechanisms as changes in the secretion of satiety hormones (Cassady et al., 2009; Kokkinos et al., 2010); lower energy intake through increased oral exposure (Bolhuis et al., 2013a; Weijzen et al., 2009) or through the number of chews per unit of food (Bolhuis et al., 2013b, 2011); decreased feelings of deprivation by enhancing and prolonging pleasurable aspects of eating (Brownwell, 2000); and changes in the encoding of the meal in memory, which in turn influences food choices in subsequent meals (Higgs & Jones, 2013; Higgs, Robinson, & Lee, 2012). Finally, the vibrotactile feedback may have made people more aware of their goal to lose weight or eat less, which may have strengthened their resolve in other nutrition-related behaviors than eating rate, such as eating healthier meals (Northcraft et al., 2011). Further research will shed light on which (combination of) underlying mechanisms causes the reduction in BMI.

An important strength of this study was the use of scientifically rigorous research and analysis methods to assess the effectiveness of technology-based feedback on eating rate and body weight in ecologically valid (real life) settings. Our study contributes to a stronger empirical base for feedback-based interventions delivered through digital technology in altering deeply engrained habitual behaviors. Despite these strengths, a few limitations should be mentioned. First, there was a drop-out level of 29% over a 15-week period. Although this level of drop-out is not uncommon for field trials, we cannot rule out the possibility
that bias from drop-out could have affected our results. The fact that half of the participants who dropped out (15% of all participants) did not receive the intervention due to technological issues (e.g., water damage to the measuring device of the fork), however, may limit the potential impact of drop-out bias on intervention outcomes. A further issue is that we were unable to address which and how many dietary consultations participants received during the study period. In addition, it is possible that their dieticians’ individual approaches differed, leading to different amounts of weight loss unrelated to our manipulations. Our randomization strategy, in which we matched participants to a computer-generated list of random numbers in order of enrollment, made sure that participants recruited by a particular dietician were randomly assigned to different conditions. Nevertheless, extra data on the specific details of the dietary treatment during the training period might have identified whether and how dietary consultation is associated with fork use and/or behavioral outcomes. Third, we could not include daily dietary or nutritional intake measures in our study, as this would have put a high burden on individuals participating in our study. Repeatedly asking participants for the type and amount of food consumed while using the fork, would have induced a demand characteristic that would have interfered with the ‘natural’ use of the fork in the training phase (Robinson, Kersbergen, Brunstrom, & Field, 2014), compromising the validity of our findings. As a result, however, we do not have an indication of the type and amount of food people ate nor do we have a reliable measurement of individuals’ satiety levels after the meal. Although the specific aim of the present study was not to examine how fork use would affect food intake and weight, further research with more exhaustive registrations of participants’ meals and satiety can help to better understand how a lower eating rate leads to weight loss.

Our findings have direct implications for the clinical management of weight control, highlighting the importance of a slow eating rate in addition to more traditional dietary instructions on what and how much to eat. The results reported here suggest that technology-based feedback provides dieticians and other care professionals with an effective intervention to slow down eating rate. This intervention, combined with other interventions aimed at healthier eating and increased physical activity, may help dieticians in supporting clients in achieving lasting weight loss, reducing physical complaints, and in the prevention and perhaps even the treatment of debilitating conditions such as diabetes type II (Ades, 2015).

All in all, this study showed that feedback from an augmented fork to decrease eating rate may be an effective tool to reduce eating rate and promote weight loss. After a short training period, people who received feedback on eating rate manage to achieve better spacing between bites, leading to less over-speed bites, longer pauses between bites, less bites per meal, and significant weight loss. Our results indicate that these effects remain after an eight-weeks pause. Further research is
Chapter 6

needed to shed light on longer-term efficacy of this intervention on eating rate and body weight.
Effects of a technology-based intervention to decelerate eating rate: a randomised controlled trial
Chapter 7: Discussion and Conclusions

Introduction

This dissertation examined the question of whether feedback delivered through digital technology changes habitual behaviour. Furthermore, we aimed to shed light on how and when this effect differs because of intrapersonal (e.g. character traits, psychological states such as motivation) or interpersonal (contextual or systemic) moderators, or because of different feedback properties. To address these questions, we performed a review of the current literature and performed studies using two existing designs for behavioural change: a physical activity tracker, and a ‘smart’ fork that gives feedback on eating rate. This chapter discusses our findings and provides an overview of our conclusions and some challenges we identified.

Conclusions from our literature review

To evaluate the effect of feedback delivered through digital technologies on habitual behaviour, we reviewed the relevant literature. A combined search in a range of scientific and design- and HCI-oriented databases, and auxiliary searches, yielded a set of 69 original papers (with a total of 72 studies) that matched our inclusion criteria: offering an analysis of the effect of digital technology that delivers tailored feedback by an external agent to provide information regarding task performance, aimed at automatic (habitual) behaviour.

We thematically classified target behaviours of the intervention, feedback technology, feedback characteristics (content (feedback sign, comparison, and level of tailoring), timing, modality, frequency, duration, data source), and the availability of visual examples of the design and provided feedback. For each intervention, we assessed the number of participants, independent and dependent variables, analysis method, results, and possible methodological concerns.

Our analysis showed strong evidence for the hypothesis that feedback disrupts habitual behaviour. 59 of 72 studies, and 13 of the 14 well-set up experimental studies with ample statistical power, show a beneficial effect of feedback on the ability to disrupt undesired habitual behaviour. This effect occurred regardless of dependent variable category, ranging from energy consumption to motor skills and physical activity. Where feedback did not lead to disruption of current behaviour, this was sometimes due to (deliberate) misunderstanding of the design’s purpose; moreover, most of the studies with null findings may very well suffer from insufficient statistical power to detect significant differences, since descriptive results in many of those studies did point towards a small positive effect of feedback on behaviour.

Our review shows that the disruptive effect of feedback on undesired habits occurs independently from modality (e.g. visual, auditory, or tactile feedback), timing, frequency, and medium (e.g. mobile phone apps, websites, or wearable devices). This may very well be the result of optimisation and iterative user testing in the design phase, which, if performed well, leads to choices for feedback modality, timing, frequency, and media that fit the target behaviour and user needs. For instance, in the case of modality, the target behaviour often rules out specific feedback modalities. In driving a car, the visual channel is more often than not occupied by keeping track of traffic. Visual feedback on driving behaviour is more often than not dangerous instead of supportive, as anyone who has ever attempted to text while driving will realise. At the dinner table, both the visual and the auditory channel are occupied, and a designed artefact which relies on visual or auditory feedback on eating behaviour needs to deal with the social practices of eating, which, for many people, has an important social function as well. Disrupting this social aspect with feedback messages on eating behaviours can be perceived as rude, and such designs are likely to be abandoned.

The hypothesis that feedback from digital technology disrupts undesired habits is supported by strong evidence. However, current literature does not (yet) provide evidence for lasting effects of this disruption on behaviour. A recent meta-analysis (Noah et al., 2018) casts doubt on this hypothesis, and shows that in general, small to non-existent effects of feedback on behaviour can be expected. Once again, this meta-analysis justifies the conclusion that much of the research that has been performed so far in this field suffers from methodological shortcomings. In a similar vein, recent trials in which feedback is used to increase physical activity report null effects for feedback alone (Cadmus-Bertram, Marcus, Patterson, Parker, & Morey, 2015; Finkelstein et al., 2016). Therefore, our review enables us to at least partially answer our first question: yes, feedback from digital technology is able to disrupt undesired habits; but whether this leads to lasting behavioural changes remains unclear. To test this hypothesis, we need more research with
higher quality research designs, data gathering, and analysis than what is currently common; be it qualitative or quantitative, or action research (such as the different flavours of research-through-design), which all have their relative merits to add to our knowledge. Furthermore, our review shows that there is as yet hardly any evidence available about what moderates sustained use of digital feedback technologies. The questions who uses these technologies, why they use it, in which circumstances and to which effect still need to be answered.

Conclusions from the Fitbit Study
Our literature review revealed that little is known about what factors (be it states, traits or context) drive uptake and sustained use of feedback interventions, and how uptake and sustained use differs for individuals across different contexts. The only evidence currently available comes from industry whitepapers (Chen, 2015; Fox & Duggan, 2013), which claim that 30–80% of users abandon the technology within the first weeks, depending on technology. To find out more about patterns in who will persist in the relevant technology long enough for the behaviour change to occur, we analysed tracker data and questionnaire results from 711 participants from four urban areas in France. The questionnaires measured a range of potential determinants of sustained use: demographic and socio-economical, psychological, health-related, goal-related, technological, user experience-related and social predictors. We determined the relative importance of all included determinants on the duration of tracker use by using machine learning analysis techniques.

The data showed a slow exponential decay in physical activity tracking, with 73.9% (526/711) of participants still tracking after 100 days and 16.0% (114/711) of participants tracking after 320 days. On average, participants used the tracker for 129 days. This decay is exponential, but slower than may be expected from what little literature exists on the topic. Most important reasons to quit tracking were technical issues such as empty batteries and broken trackers or lost trackers (21.5% of all respondents of our third questionnaire, 130/601). Major determinants of tracking duration were age (the under 25 kept up tracking less long than older participants) and user experience-related factors (those who liked the design and user interface of the Fitbit more and found it easier to use, tracked longer than those who liked it less and found it more challenging). Other, smaller determinants were mobile phone type (iPhone less than others), household type (single parents less than others), perceived effect of the Fitbit tracker, and goal-related factors (having ‘adjacent’ goals such as healthy eating and quitting smoking decreased Fitbit use, when compared to ‘central’ goals such as increasing activity). Interestingly, many determinants had a smaller contribution to sustained use than may be expected from literature, or no effect at all. Perhaps this means that in real
Discussion and conclusions

Life, determinants such as education, character traits, income, and profession play a much smaller role than in isolated lab conditions.

Intuitively, it may not seem surprising that user experience, aesthetic preferences, and ease of use matter to sustained use, especially to professionals from practice, but in the scientific community, many are relatively unaware of their importance and put more faith in underlying general working mechanisms and neglect user experience design (Hermsen, Van der Lugt, Mulder, & Renes, 2016), which may lead to clunky designs.

All in all, our Fitbit study shows that people who start using an activity tracker, can be expected to use their tracker for a longer period than previous literature predicted. Furthermore, seamless technology and a positive user experience, combined with a fit between technology and tracking goals, lead to sustained use; in this study, their effect is greater than that of many of the determinants currently favoured by psychological and digital health research.

Conclusions from our ‘smart’ fork studies

To contribute to the existing knowledge on whether feedback through digital technology can change undesired habits, both directly and in the long run, we performed three studies that evaluate the acceptance and efficacy of a design to slow down eating rate. Eating rate is a deeply engrained habitual behaviour, which is strongly associated with stomach disorders and overweight (Robinson, Almiron-Roig, et al., 2014). The latter causes a range of debilitating health issues, such as diabetes type II and some forms of cancer (Berenson, 2012). Because of its deeply automatic nature, eating rate is put-near impossible to change by will alone. The research presented here tests the hypothesis that real-time feedback from a ‘smart’ fork, equipped with sensors and actuators, enables people to durably change their eating rate. The deeply engrained, highly automatic nature and detrimental effects of eating rate make it an ideal candidate for research into the potential of technological solutions that provide feedback on behaviours. Non-technical solutions (e.g. mindfulness) are destined to fail because of the amount of willpower needed to keep up the desired behaviour, and many technical solutions fall short because of their lack of fit with the social practice of eating: an oral restriction device (McGee et al., 2012) leaves users with speech difficulties; a scale-based device (Zandian et al., 2009) is too conspicuous for use in many public settings or fails to engage (Hamilton-Shield et al., 2014). The smart fork is an example of a technology that is less intrusive than most existing solutions, and theoretically usable in social contexts. It may provide a feasible solution to change a deeply engrained behaviour.

Our first aim was to assess the acceptability of the smart fork in everyday contexts. To evaluate the usability and acceptance of this fork, which is available on the market under the name 10sFork, we asked 11 participants to eat a single meal
with the fork in our laboratory, and then take the fork home for three days and use it as much as possible. After the laboratory meal and upon returning the fork, we interviewed the participants. The fork proved an acceptable tool: users reported enhanced awareness of their eating rate, both in the laboratory and at home, and felt comfortable using the fork in social settings. However, none of the participants felt the fork was ‘for them’, even though they did recognise the need to slow down their eating rate. This mismatch between self-perceived target group membership and the solution may be an issue affecting acceptance of the fork as an intervention for healthy eating in real life. When it comes to user experience, all participants thought the fork was comfortable and easy to use. Biggest perceived issues were the fact that the fork does not take bite size into account, and the fact that some foods triggered false-positives: feedback where no bites were taken.

Acceptance and usability, as preconditions for effective behaviour change, are evaluated for most feedback interventions, but hardly any digital health tool undergoes rigorous testing of its efficacy. For instance, of the 400 physical activity apps currently available in the diverse app stores for mobile phones, only 12 had peer-reviewed study in any shape or form connected to them (Bondaronek, Alkhaldi, Slee, Hamilton, & Murray, 2018); consumer health wearables show a similar lack of scrutiny (Piwek, Ellis, Andrews, & Joinson, 2016). To see whether the ‘smart’ fork actually succeeds in decelerating participants’ eating, we first looked at whether the fork is able of slowing down fast eaters in a single sitting. Subsequently, we looked at the effect of real-life use of the fork.

To test the effect of the fork on eating rate in a single meal, we invited 114 self-reported fast eaters to our lab. They were randomly assigned to a feedback condition, in which they received vibrotactile feedback from their fork when eating too fast (i.e. taking more than one bite per 10 seconds), or a non-feedback condition, where they ate with the fork without feedback. To control for demand characteristics, we told all participants about the importance of eating slowly, and that the fork would record their eating speed. Participants in the feedback condition ate at a slower rate (fewer bites per minute) and had a higher success ratio (percentage of bites taken after a pause of at least 10 seconds) than did those without feedback. A slower eating rate and higher success ratio, however, did not lead to a significant reduction in the amount of food consumed or level of perceived satiation. This may have to do with the artificial setting of the meal; uncertainty about norms in a social setting are known to cause people to ‘revert’ to generally accepted ideas of portion size (Higgs et al., 2012). Alternatively, a slower eating rate may take more meals to start having an effect on the amount of food we eat.

Finally, we performed a field study, to learn more about the effect of using the fork in everyday life. We enlisted 163 participants, all self-reported fast eaters. To make sure all participants were well motivated to change their eating rate, we
invited only participants currently under treatment of a dietician for complaints related with their eating rate, such as overweight and stomach complaints. All participants used the fork for one week without feedback to establish their baseline eating rate. Then they were randomly assigned to one of three conditions: eating as many meals with the fork as possible for four weeks, without feedback (control condition: NFC group); same, but with vibrotactile feedback (VFC group); and same, but with vibrotactile feedback and access to an online dashboard that provides retrospective feedback on eating rate (VFC+ group). After this four-week training period, they once again ate with the fork without feedback for a week, to establish the effect of the training. This one-week measurement was repeated eight weeks later.

At post-measurement, participants in both intervention groups (VFC and VFC+) had a significantly higher success ratio (percentage of bites taken after a pause of at least 10 seconds) than those in the control group (NFC). These effects persisted after an eight-week washout period. Eating rate (bites per minute) only changed significantly for the VFC group, and only directly after the training period, which suggests that the effect of the feedback on eating rate mostly consisted of better spacing between bites, with less ‘fast’ bites and more ‘slow’ bites within a meal. Receiving vibrotactile feedback also resulted in a significant loss of weight directly after the training for both experimental conditions (NFC: no weight loss, VFC on average 0.56 BMI points, VFC+ on average 0.62 BMI points), with no rebound after eight weeks.

All in all, these ‘smart’ fork studies show that it is indeed possible to disrupt and durably change a deeply engrained behaviour through the provision of feedback alone. Moreover, the feedback from the ‘smart’ fork not only affected the outcome measure the feedback was aimed at, but also had an effect on users’ weight. This finding is not in line with some of the findings in our literature review and other recent publications (most notably Noah et al., 2018), which provide evidence for the disruption of undesired habits, but not necessarily for lasting effect of feedback on habitual behaviour. In the following sections, we will discuss the robustness and generalisability of our findings.

**Durable, sizeable, generalisable results?**

The results of our studies suggest that, in general, feedback from digital technology indeed has the potential to change undesired behaviours, and that when people use digital technology to provide them with feedback, they will show greater engagement with the technology than we could assume from previous literature. However, questions about the durability, the size and the generalisability of
the effect of feedback on undesired habits still remain unanswered. A first question is the duration of the feedback effect. In many behaviours and associated outcome measures, rebounds can be expected at some point (e.g. in weight loss: Curioni & Lourenço, 2005). In most studies, long-term effects are defined as occurring after at least twelve months (Johns, Hartmann-Boyce, Jebb, & Aveyard, 2014). In those studies where we measured behavioural change, we did not have the opportunity to evaluate the effects of the feedback for more than eight weeks after the end of training. Further research is needed to shed light on the durability of the effect of the fork, and whether rebound effects occur.

A second question concerns the size of the observed effect. The effect on success ratio (the proportion of 'slow' bites) was large, but the effect of the feedback from the fork on weight status was more moderate; on average, participants lost between 1.2 and 1.5 kg due to using the fork directly after the training period, and a further 1.0 to 1.8 kg during the eight-week follow-up period, whereas participants in the control condition did not lose weight. A weight loss of two to three kilograms on an average weight of 93 kilograms is a significant result, but recent findings (e.g. Cadmus-Bertram et al., 2015; Finkelstein et al., 2016; Noah et al., 2018; Patel, Asch, & Volpp, 2015) suggest that, for most behaviours, feedback on its own is not enough to durably change behaviour in such a way that no further change is necessary. For instance, in the case of our activity tracker research, would we have had the opportunity to contrast our dataset to a similar set of people who did not receive feedback, there would have been a large chance that we would not find an effect of tracker use on behaviour at all, let alone a large effect.

Bigger effects are achieved by interventions which do not (just) rely on feedback, but also provide additional behaviour change techniques – such as digital coaching (Wijsman et al., 2013), social support (Finkelstein et al., 2016), rewards (Mitchell et al., 2018), action perspective (White et al., 2017), and gamification (Althoff, White, & Horvitz, 2016). This evidence calls for a more holistic approach, in which feedback is delivered as part of a more comprehensive intervention, which also contains other behaviour change techniques aimed at enhanced engagement and clear, feasible goals.

A third question concerns the generalisability of the effects of vibrotactile feedback on eating rate to other behaviours and different forms of feedback. As we have seen before, our findings contrast recent meta-analytical results (Noah et al., 2018), in which feedback from digital technologies have only limited effect on habitual behaviour, and some results of our own literature review, which also casts doubt on the potential efficacy of such feedback. It could very well be that the smart fork was an exceptional case in terms of efficacy, where both the behaviour and the technology had properties which led to an exceptional fit which cannot be expected in other contexts. Evidence shows that feedback on behaviour should be timely,
personal, and actionable (Schembre et al., 2018). The feedback from the smart fork satisfied all three requirements. Users received feedback on eating rate if and only if they ate too fast, at the precise moment when the behaviour occurred, in the form of a slight buzz. As long as people ate slowly, they received no feedback. This setup encouraged independent and conscious scrutiny of eating rate.

When not all three requirements for effective feedback (timely, personal, actionable) are met, the chances of effect are potentially much slimmer. The feedback provided by the Fitbit in our activity tracker study is personal, but only partially actionable because not everybody will, in every circumstance, be capable of increasing their physical activity. Furthermore, the step count is aimed at daily appraisal; it resets at midnight. This invites retrospective evaluation of your performance through the possibility of comparing different days, but this also means that evaluation does not necessarily take place when the (lack of) behaviour occurs. This may (partially) explain why feedback when given by itself seems insufficient to change physical activity (Cadmus-Bertram et al., 2015; Finkelstein et al., 2016). Furthermore, contrary to findings in our literature review, the way the feedback is provided could very well affect performance in the long run, even if it helps achieve initial successes. For instance, the 10sFork has the possibility to flash a green light when a new bite is allowed. Evidence (Stawarz, Cox, & Blandford, 2014, 2015) shows that this kind of feedback, where people receive a cue that ‘allows’ them to perform a behaviour, creates a dependency between the new cue and the desired behaviour. People rely on the intervention to interpret the behaviour and choose the action, and when the intervention is taken away, the effect disappears. Fortunately, the default setting of the fork, which shows a red light when people eat too fast, does not have this drawback.

Concurrent feedback, while the behaviour occurs, is possible because of automatic measurement of the target behaviour. Both the Fitbit activity tracker and the 10sFork benefit from the availability of very small, low-cost movement sensors and actuators to measure the target behaviours. Furthermore, newly developed algorithms help determining whether a sensed movement constitutes an instance of the selected target behaviour or a false positive. For many other behaviours, reliable, automatic measurement of performance is (as yet) impossible; for even more behaviours, it is not yet possible to provide reliable real-time feedback. This lack of automaticity may very well be the cause of the lack of effect found in many trials in which feedback technologies have been tested. For instance, the automatic analysis of nutrition is not yet feasible (but see Tseng, Napier, Garbarini, Kaplan, & Omenetto, 2018 for a project which shows potential), even though the first products that claim to do so have already appeared online (e.g. Fitly, 2017). To obtain feedback on nutritional content, the user still has to painstakingly provide their own data. This is very likely to have a detrimental effect on sustained use (e.g.
in Hermsen & Frost, 2018). Furthermore, self-report of nutritional content is often unreliable (Archer, Hand, & Blair, 2013; Dhurandhar et al., 2015; Ioannidis, 2013; Mitka, 2013), which affects the quality of the provided feedback. Similarly, technical solutions that reliably recognise human emotions are infeasible and will be for the near future.

However, the number of behaviours and outcome measures that can now be reliably measured through passive sensing rose steeply in the past few years. The increased potential for the automatic collection of data, which allows interventions to reach areas of life which were previously inaccessible, has led to a similarly steep increase in the number of designs that provide feedback on behaviour. This has been most visible in the field of physical activity, where the availability of a range of low-cost trackers has led to an exponential growth of tracker-based studies and interventions (Müller et al., 2018).

Even when it is feasible to automatically evaluate behaviour, real-time feedback is not always possible or useful. As we have argued before, meaningful real-time feedback on (for instance) physical activity is difficult, because it is hard to judge whether a person is currently able to be more physically active, or if this activity should be performed at a later moment. A relatively new and promising development in the automatic evaluation of complex behaviours, is the application of machine learning principles to signal events that warrant feedback. There are designs that predict influenza (Barlacchi, Perentis, Mehrotra, Musolesi, & Lepri, 2017) and depression (Mehrotra, Hendley, & Musolesi, 2016) from human activity patterns, and it is now possible to reliably predict when people who just quit smoking are in danger to start again (Naughton et al., 2016). These automatic measurements make it possible to provide feedback on complex health issues and events as they occur, whereas until now, feedback could only be given when it was already too late to act upon it. In the case of activity tracking, for instance, this may mean that the device provides cues to increase physical activity at times when evaluations of normal behaviour patterns predict a greater physical activity than currently measured. Such developments, both in measuring behaviours, and in delivering real-time feedback, broaden the scope of potential designed solutions that provide us with feedback on our behaviour. However, many behaviours and human practices are (and will be for quite some time) too complicated to measure.

**Engagement with feedback interventions**
The effect of the ‘smart’ fork on eating rate and weight status, albeit small, suggests the potential of digital feedback technology to change undesired habits that until recently were thought near impossible to change. This confidence is further enhanced by the findings of our Fitbit study: people tend to use their tracker longer than could be expected from current literature. But new challenges also
emerged. One of the most important threats to durable behaviour change is a lack of engagement with the behaviour and with the intervention (Perski, Blandford, West, & Michie, 2016). Evidence shows that lasting engagement with the intervention is essential to behaviour change (Couper et al., 2010; Donkin et al., 2011; Funk et al., 2010), since people need to use an intervention for a certain period for behaviour change to set in. But even starting to use an intervention needs a certain level of engagement. Both the fork studies and the physical activity tracker study describe situations in which people have already volunteered to use the intervention; deciding if you need an intervention for your health, and then voluntarily starting to use it, is a wholly different question.

In general, it is hard to get people to accept the gravity of a problem, and even harder to convince them to accept solution as being specifically suited for their own situation. People as a rule do not like to engage with detrimental health conditions (Swinkels et al., 2018). The motivation to monitor and change our behaviour needs to be very high to even download an ehealth app (Hermsen & Frost, 2018), people need to have a certain amount of perceived self-efficacy, and people need to be aware of both the problematic behaviour and the severity of its consequences, to start using a digital feedback technology. Both in our user experience evaluation study and in our lab study, participants did not particularly feel the need to change their eating rate. Even after receiving information about the detrimental long-term effects of eating too fast on our health, they did not feel motivated to slow down.

A similar effect of engagement on behaviour change can be seen in our Fitbit-study, where sustained use of the intervention was highly correlated with the measured physical activity on each day of use. Our Fitbit-study provided further evidence for the importance of engagement for feedback efficacy in the relative importance of fitting goals for sustained use. Intrinsic goals are highly correlated with, and are known to predict, self-determined motivation (Gillison, Standage, & Skevington, 2006; Wilson, Mack, & Grattan, 2008). A lack of motivation is a threat to the expected efficacy of these feedback interventions in real life.

One approach to enhance engagement is to add features to the intervention that provide an incentive to start using it, and to use this incentive as a driver for sustained use. These features can take the form of rewards, game elements, narratives, social and emotional support features, and many others. This top-down strategy can persuade potential users to adopt the design to support their behaviour change. Unfortunately, little agreement exists among users and experts on what features work (Herbec et al., 2018; Perski, Baretta, Blandford, West, & Michie, 2018). The effect of engagement-enhancing features may differ greatly for different people across different context, and a mismatch between an attempt to create greater engagement and user preferences can easily lead to abandonment.
Chapter 7

Introducing game elements, for instance, will greatly enhance engagement for some users, but is likely to drive away others.

A more fruitful, but perhaps more challenging approach towards a greater match between user needs and feedback interventions lies in the concept of lived and personal informatics (Epstein et al., 2015; Li et al., 2011; Rooksby et al., 2014). Personal informatics tools aim to support reflection on personal data to generate insights for self-improvement for health, productivity or well-being (Li et al., 2011). This is a bottom-up approach, that encourages users to actively obtain (self) insights by examining their data and testing self-generated hypotheses, and subsequently change their behaviour based on these insights. The premise of personal informatics approaches is that different people in different context have different needs for data to change their behaviour, and therefore different hypotheses they would like to test. People’s goals and needs shift over time, and so do reasons to use an activity tracker: these can range from using a tracker to decide whether physical activity tracking is something you would want to do, tracking your behaviour to find out what your current physical activity is, using your tracked data to diagnose symptoms not directly related to the amount of steps taken, such as general fatigue, using the data to direct behaviour change or to interact with others all asks different questions, to simply using a tracker because of an interest in technology. These very different reasons to use an activity tracker all have different ramifications for the design of feedback interventions.

Interventions based on personal informatics help their ‘users’ to put their health concerns in more concrete terms, help to prioritise hypothesis testing, and support sustained behavioural change. The difficulty is the tailoring to personal needs; some progress has been made in recent years to identify patterns in user needs (Epstein et al., 2015; Rooksby et al., 2014). More research into these patterns, and into tailored ways to concretise intrinsic motivations for healthy behaviour change, evoke positive emotions, and offer rewarding interaction during the use of feedback interventions, will enable us to design more engaging feedback interventions that align with the needs and possibilities of their users.

To track or not to track: appropriateness and privacy

Machine learning techniques and other forms of automatic measurement of behavioural data have their advantages, but they also come at a cost. The fact that sensors and actuators make real-time feedback on behaviour possible, does not necessarily mean that we should encourage the use of this kind of feedback. A couple of critical notes on the ethical desirability of feedback interventions are appropriate. Some scholars have described self-tracking in Foucaultian terms, where subjects willingly regulate, govern, and optimise themselves (e.g. Whitson, 2015). There is indeed a fine line between beneficial self-regulation through feedback, and the use
Discussion and conclusions

of automatically generated behavioural data to subject people to standardisation and regulation. For instance, many people do not want to be confronted with their health, let alone improve their behaviour (Swinkels et al., 2018). People have the right to autonomy (up to the point where their freedom starts threatening the freedom of others), even if this means that they make choices that are detrimental for their health. We should be vigilant towards threats to peoples’ freedom to deviate from the norm, as self-tracking practices are slowly becoming incorporated into different areas of social life. Educational institutions, workplaces, and even insurance companies, are investigating ways to encourage people to adopt digital feedback technologies to enhance productivity, health, and information gathering. This imposed or exploited tracking (Lupton, 2014) threatens our autonomy – failure to participate in wellness programs at work may already lead to higher health insurance premiums, as is happening in some workplaces in the United States (Olson, 2014). A choice to take part in a feedback intervention should therefore be voluntary. Deciding not to take part should be without punishment in any form whatsoever, be it financially or through access to health care.

We should also be cautious of feedback interventions that lead to greater health disparities. People experiencing socioeconomic disadvantages face a range of challenges that can substantially hinder efforts to adopt healthy behaviours (Ball, 2015). For these groups, feedback on their behaviour alone is hardly ever sufficient to drive behavioural change. In circumstances where high-SES groups may only need information on occurrence and consequences of their behaviour, low-SES groups experience additional barriers, often economic or systemic in nature, that can only be reduced by broad, communal approaches. Furthermore, people in socioeconomically disadvantaged positions often need additional behaviour change techniques that provide social and emotional support to successfully negotiate a behaviour change process (e.g. in Brown et al., 2014). Feedback interventions that are only effective for high-SES groups and do not contribute to greater health equality are ethically and politically undesirable. As a consequence, we need a much larger body of research into how low-SES groups and other groups with limited health literacy can benefit from feedback interventions, and how these interventions should be enhanced to better serve these groups.

An essential part of the design process of any feedback intervention should be designing for the largest possible privacy and trust. In intervention design, privacy and trust hardly ever seem an issue, but in real life and in the news, they have received a lot of attention recently (e.g. because of the Facebook - Cambridge Analytica crisis, Lin, 2018). However, feedback interventions do give rise to privacy concerns. In order to give feedback, most products rely on data analysis that takes place on the vendor’s servers, and visualisation of feedback through online and mobile applications. Who owns the data that is generated by measuring
your behaviour? Who guarantees that this data remains within the closed loop of measurement - analysis - feedback - measurement, and does not get transferred on to third parties? Is your data accessible? TherosFork in the vibrating fork project registers each and every bite with a unique time stamp. This data is then used for direct, vibrotactile feedback, and for retrospective visualisations of your eating rate patterns. This data set was until recently unavailable for users of the fork. Similarly, many activity trackers such as the Fitbit will tell you how many steps you have taken, will show you historic trends in your activity and how you compare to others, but the entire data set with every registered step remains unavailable, stored in the vendor’s server park for who knows what use. In the case of the Fitbit, a recent report (Forbrukerrådet, 2016) shows that the end user licence agreement does not tell users whether their data will be shared with third parties and whether and how long their data will be retained after the user deletes them. Furthermore, the Fitbit appears to collect more data than strictly necessary to provide feedback on physical activity, and the producer of the tracker does not announce changes in the end user licence agreement.

When designing feedback interventions, we need to take user privacy into account, and make the data available to those who rightfully own them (i.e. the users) in a safe and transparent manner. It may take a change in legislation, such as the new GDPR guidelines in the European Union for ethically viable data management to occur, but this does not relieve feedback intervention designers of the burden of carefully thinking through the consequences for privacy their design has.

To summarise, both in academia and practice, when developing interventions that use feedback as a driver for behaviour change, the focus should lie on designing and researching technology that allows people to answer their own questions through personal informatics and restrict themselves to an ethical dimension in which private (gathering data for one’s own purposes only) and communal (voluntary sharing among trusted peers) tracking is encouraged, and pushed, imposed, or exploited tracking is avoided. Such feedback interventions should be embedded in broader approaches that provide the kind of support and focus people with little health literacy need. Both the Fitbit and the smart fork contain some of the ingredients necessary to make them a tool towards greater health equality. However, they also provide the possibility for unethical use, towards standardisation and surveillance, depending on how they are deployed. The Fitbit per se does not provide a useful way to support disadvantaged groups; providing feedback without further behaviour change techniques aimed at embedding in community approaches or providing emotional and social support sooner enlarge health disparities than reduce them. The smart fork may be more suitable to support greater health equality, but as yet it is uncertain whether the
feedback from the fork in itself provides the kind of support needed by lower health literacy groups.

One final issue remains. Unfortunately, although research presented in this paper shows the potential efficacy of designs that provide feedback, we are currently still far from understanding the precise cases in which feedback through digital technology can sustainably change behaviour, and from finding out what works for whom in what contexts. Our literature review shows that this state of affairs is not helped by current methodological standards and reporting traditions in fields where much of the research takes place: design research and HCI. These are insufficient to generate generalisable knowledge about the efficacy of our designs for behavioural change. The field of digital health (ehealth, mhealth, etcetera) normally has more rigorous scrutiny, but in this field, again, our literature review shows that only a few interventions are tested for long-term efficacy. An additional problem in the field of digital health is the speed at which research is conducted. The gold standard for evaluating interventions, the randomised controlled trial, is too costly and, above all, too slow (Ben-Zeev et al., 2015; Riley, Glasgow, Etheredge, & Abernethy, 2013) for the rapidly evaluating technologies that provide feedback on behaviour. Current standards require interventions to remain stable during testing, which often means that the technology on which the intervention is based is obsolete by the time the results of the trial have passed peer review. It is of utmost importance to develop more ‘efficient’ (fast, responsive (Riley et al., 2013), and agile (Hekler et al., 2016)) research methodologies to evaluate interventions. Furthermore, the randomised controlled trial does not provide answers on what works for whom in which contexts, but only justifies more general conclusions. This does not fit the more tailored approach we have been advocating in this text. We should be putting more effort in employment and further development of rapid research methods such as single case designs (McCallum, Rooksby, & Gray, 2018) to evaluate mobile health technologies, and more thorough reporting standards of the research involved in the design process and its iterations. Only then can we make more generalised conclusions about what kind of feedback intervention works for whom in which context.

**Conclusion**

This thesis examines two questions: does feedback through digital technology have an effect on undesired habitual behaviour, and what determinants and feedback properties enhance the efficacy of the feedback? We have seen that feedback from digital technology can disrupt the automatic cue-response-pair of habitual behaviour, which makes that behaviour available for conscious scrutiny.
All this does not necessarily mean that we can expect the technology and the feedback itself to lead to behaviour change. Current evidence allows us to see feedback through digital technology as a vehicle for behaviour change, but not (yet) as a driver (Patel et al., 2015). Those who are already willing (and capable) to change their behaviour, can use feedback technology as a means to do so, whereas for those who are unwilling (or lack the capability) to change their behaviour, feedback technology in itself will not be enough. The ‘smart’ fork project, however, shows that in cases where it is possible to provide timely, personal, and actionable feedback, derived from automatically gathered data and delivered through an appropriate channel, feedback on undesired habits can be a very effective behaviour change technique. Our research shows that digital technology, in the right circumstances, is indeed capable of changing undesired habits that were previously out of reach.

Our second question: which determinants or feedback properties enhance efficacy, proved harder to answer, but here as well, our research provides starting points for both further research into feedback interventions, and for their design. The Fitbit project showed that, once people have adopted a feedback technology, they will use it quite extensively, as long as the technology holds up. Our research with both the Fitbit activity tracker and the ‘smart’ fork shows that user experience and engagement with the intervention play an important role. Challenges lie in developing interventions that closely match the goals of the intended users, and in keeping users engaged long enough for behaviour change to occur. Unfortunately, people are not necessarily adequately motivated to take up a technology that provides the feedback as their vehicle towards behaviour change. Recent evidence suggests that for most people, feedback in itself is insufficient to provide enough motivation to warrant uptake and use of digital feedback technologies. Broader, more holistic interventions, with feedback as one of the central behaviour change techniques, are needed to persuade people to start using feedback technology, and to support them in their behaviour change processes.

This thesis ends with defining a range of challenges for intervention design practice. Firstly, we need to design products or services in such a way that people understand that they are the target group, and in a way that motivates uptake, without discouraging users by scaring them off or triggering cognitive dissonance reduction. Secondly, we need to work on new methods to evaluate our designs properly, be it through qualitative, quantitative, or action research. Thirdly, we must develop best practices for using feedback for behavioural change in a way that satisfactorily address privacy concerns. We must work on solutions that provide open, usable data for users which remain inaccessible to anybody else. And finally, in designing our interventions, we must focus more on inclusiveness, making sure
our designs support not only those with high health literacy, but (and foremost) those who find themselves in less privileged circumstances.

To do so, we must embrace a broader view on the accumulation of knowledge. The projects described in this thesis are all the result of broad, interdisciplinary cooperation, not only between scientists with very diverse backgrounds and research interests, but also between academia and practice. The data in the Fitbit project were gathered as a result of a publicity campaign by a private communications agency, who then sought cooperation with researchers from academia to evaluate the data. The ‘smart’ fork project was supported by a grant scheme that aims to further private-public cooperation in science. Such diverse interdisciplinary approaches are needed to transcend the limitations that each field by itself is subjected to. We hope the research projects in this thesis inspire others to embrace this broader view.
References


References


Chapter 8


References

Faruqui, A., Harris, D., & Hledik, R. (2010). Unlocking the €53 billion savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU’s smart grid investment. Energy Policy, 38(10), 6222–6231. http://doi.org/10.1016/j.enpol.2010.06.010


Chapter 8


References


References


Chapter 8


Lin, Y. Y. (2018). #DeleteFacebook is still feeding the beast — but there are ways to overcome surveillance capitalism. Retrieved April 11, 2018, from http://www.webcitation.org/6yboh59AU

References


Chapter 8


References


Chapter 8


Chapter 8


Supplementary Materials

Appendix 1: Overview of included studies in our Literature Review
https://osf.io/zgjvd/

Appendix 2: All questionnaire items of the Fitbit study, with response scales, variables in which they were used, transformations, and validity evaluation.
https://osf.io/bz62n/

Appendix 3: Principal Component Analysis of items relating to User Experience
https://osf.io/9uepj/
### Summary

Habitual behaviour is often hard to change because of a lack of self-monitoring skills. Digital technologies offer an unprecedented chance to facilitate self-monitoring by delivering feedback on undesired habitual behaviour. A broad range of feedback solutions are currently available: wearable activity trackers give us feedback on whether we walk or sleep enough; smart devices track our eating habits; an app can warn us about situations in which we are likely to smoke a cigarette, and a growing number of devices tell us (and others) what emotions we experience in cases where we are unable to do so ourselves. Unfortunately, there has been relatively little research into whether all this feedback on health behaviour is as effective as we implicitly presume.

This thesis contributes to answering the question whether feedback through digital technology is effective to change habitual behaviour. To do so, Chapter 2 of this thesis provides a review of the current literature and an analysis the results of 72 recent studies in which feedback from digital technology attempted to disrupt and change undesired habits. A vast majority of these studies found that feedback through digital technology is an effective way to disrupt habits, regardless of target behaviour or feedback technology used. However, methodological issues limit our confidence in the findings of all but 14 of the 50 studies with quantitative measurements in this review; more well-designed research into the efficacy of feedback in disrupting habitual behaviour can increase our confidence in these findings. In our review, only 4 studies tested for (and only 3 of those 4 found) sustained habit change, so the current scientific state of the art is not sufficiently developed to draw a conclusion about whether feedback from digital technology can durably change detrimental habits. Furthermore, it remains unclear how feedback from digital technology is moderated by receiver states and traits, as well as feedback characteristics such as feedback sign, comparison, tailoring, modality, frequency, timing and duration.

Based on the results of this review, this thesis zooms in on two urgent research questions: Is feedback through digital technology an effective way to sustainably change habitual behaviour, and is feedback through digital technology effective for each user in every context (or are there intrapersonal (e.g. character traits,
psychological states such as motivation) or interpersonal (contextual or systemic) moderators? To do so, the thesis presents an evaluation of two existing interventions for behaviour change that provide feedback on undesired habits. Both interventions (or its successors) are currently available in the consumer marketplace. The first intervention, a wearable activity tracker, can provide insights into which intrapersonal or interpersonal determinants increase the chances of sustained engagement with the intervention. The second product, a ‘smart’ fork that measures eating rate and gives feedback when its user eats too fast, can provide insights into whether feedback can durably change a deeply engrained, rigorous detrimental habit.

In Chapter 3, I report a longitudinal study into potential determinants of the sustained use of a wearable activity tracker. Feedback from activity trackers has the potential to encourage daily physical activity and decrease detrimental sedentary habits. To date, little research is available on the natural development of adherence to feedback technology, in this case: activity trackers, or on potential factors that predict which users manage to keep using their devices during the first year (and thereby increasing the chance of healthy behaviour change) and which users discontinue using their devices after a short time. The aim of this study was to identify the determinants for sustained use in the first year after purchase. Specifically, we look at the relative importance of demographic and socioeconomic, psychological, health-related, goal-related, technological, user experience–related, and social predictors of feedback device use. Furthermore, this study tests the effect of these predictors on physical activity.

A total of 711 participants from four urban areas in France received an activity tracker (Fitbit Zip) and gave permission to use their logged data. Participants filled out three Web-based questionnaires: at start, after 98 days, and after 232 days to measure the aforementioned determinants. Furthermore, for each participant, we collected activity data tracked by their Fitbit tracker for 320 days. We determined the relative importance of all included predictors by using Random Forest, a machine learning analysis technique. The data showed a slow exponential decay in Fitbit use, with 73.9% of participants still tracking after 100 days and 16.0% of participants tracking after 320 days. On average, participants used the tracker for 129 days. Most important reasons to quit tracking were technical issues such as empty batteries and broken trackers or lost trackers. Random Forest analysis of predictors revealed that the most influential determinants were age, user experience–related factors, mobile phone type, household type, perceived effect of the Fitbit tracker, and goal-related factors.

Chapter 4 introduces the 10sFork, which provides feedback to raise awareness of eating rate in order to help people eat more slowly. It records behaviour and provides real-time haptic feedback on individual eating rates. Eating rate is a
basic determinant of appetite regulation, as people who eat more slowly feel sated earlier and eat less. As a result, fast eating rate may contribute to overeating and weight gain, and also to a range of debilitating conditions such as diabetes II, gastro-intestinal disease and some types of cancer. Unfortunately, without assistance, eating rate is difficult to modify due to its highly automatic nature. This chapter reports an evaluation of the 10sFork for usability. Eleven participants (three male, eight female) used the fork both in a laboratory setting and at home. All participants indicated having high eating rates. We interviewed them on perceived efficacy, acceptability, comfort, accuracy, motivation, and sustained use of the fork. Participants feel the 10sFork is an acceptable tool to decelerate their eating rate. The fork is generally seen as comfortable and sufficiently accurate. Participants were more aware of their eating rate but did not always feel this awareness led to (perceived) behaviour change. The vibrotactile feedback worked as expected, but the visual feedback largely remained unnoticed. Sustained motivation to use the fork was limited because participants did not see themselves as the product’s target group.

In Chapter 5, I describe the results of a laboratory study on the effect of eating with the 10sFork in a single meal setting. A total of 114 participants were randomly assigned to a Feedback Condition (FC), in which they received vibrotactile feedback from the 10sFork when eating too fast (i.e. taking more than one bite per 10 seconds), or a Non-Feedback Condition (NFC) in which they ate with the fork but without feedback. Participants in the FC took fewer bites per minute than did those in the NFC. Participants in the FC also had a higher success ratio, indicating that they had significantly more bites outside the designated time interval of 10 seconds than did participants in the NFC. A slower eating rate, however, did not lead to a significant reduction in the amount of food consumed or level of satiation. These findings indicate that real-time vibrotactile feedback delivered through an augmented fork is capable of reducing eating rate.

The long-term effectiveness of this form of feedback on satiation and food consumption is the subject of Chapter 6. This chapter reports a field study in which we assessed the effects of eating with the 10sFork on participants’ eating rate and body weight over a 15-week period. To do so, we conducted a three-armed parallel group randomised controlled trial. A total of 141 participants with overweight or obesity were randomised to either one of two intervention groups (VFC, VFC+) or a control group (NFC). In VFC, participants received direct vibrotactile feedback from an augmented fork when eating too fast during a four-week training period. In VFC+, participants received the same vibrotactile feedback, but also had access to an online web portal with retrospective visual feedback on eating rate. In NFC, participants ate with the augmented fork without any form of feedback. Eating rate (i.e., success ratio (the percentage of bites with a sufficiently long pause between
them) and bite rate) and body weight were measured at baseline (T1), directly after the four-week training period (T2) and at a follow-up after eight weeks (T3). Participants in both intervention groups had a significantly higher success ratio than those in the control group directly after the intervention. This effect persisted after an eight-week break. Bite rate only changed significantly directly after the intervention for those in VFC. Participants in both intervention groups lost significantly more weight than those in the control group after the intervention with no rebound after eight weeks. This study showed that the use of an augmented fork to decrease eating rate may be an effective tool to reduce eating rate and promote weight loss.

In Chapter 7, I discuss the findings presented in this thesis. All in all, this thesis shows that feedback from digital technology indeed has the potential to durably change (at least some kinds of) undesired behaviours. Furthermore, this thesis shows that when people use digital technology to provide them with feedback, they will show greater engagement with the technology than we could assume from previous literature.
Acknowledgements

As you set out for Ithaka
hope the voyage is a long one,
full of adventure, full of discovery.

Laistrygonians and Cyclops,
angry Poseidon—don’t be afraid of them:
you’ll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.

Laistrygonians and Cyclops,
wild Poseidon—you won’t encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.

C.P. Cavafy, 1911
Well, the journey has definitely been long and eventful, and I enjoyed (almost) every single step of the way. Fortunately, many people accompanied me on my journey and without them, I would never have reached this proverbial Ithaka. I would like to take this opportunity to thank all of you:

First of all, Hogeschool Utrecht for offering me the opportunity to apply for a doctoral grant. In my work, I have never felt so free to pursue my own ideas as during the last five years. My promotor, Peter, for your supportive, relaxed, and insightful supervision, which is quite rare among supervisors (or so I am told), and for putting up with my freewheeling and sometimes somewhat overly autonomous approach to the whole process. My co-promotor Jeana, for your insightful ideas and for taking up the ungrateful task of intercepting my attempts to add new, self-constructed words to the English dictionary. My co-promotor and supervisor at the Utrecht University of Applied Sciences, Reint Jan Renes, for your enthusiasm, unquestioning support, and for creating excellent research opportunities. Furthermore, I would like to thank all my colleagues at the research group Crossmedial Communication in the Public Domain for being such a great bunch.

My co-authors and all the others who worked with me on the projects that led to the chapters in this thesis, Jonas Moons, Carine Wiekens, Martijn de Groot, Monica Mars, Eric Robinson, Suzanne Higgs, Jacques Lépine, Jean-Baptiste Schmauch, Mirna Klaiber, Jantine Juliana, and Gjalt-Jorn Peters, for all your hard work and and insights. It would never have happened without you.

A special word of thanks for Roel Hermans, for everything you taught me about academic life, but also for the (ongoing) conversations about the difficulties of building a career on the cutting edge of science and practice, and for winding down with champions league football and a beer in a pub after a busy conference day.

My family and my friends, my parents, Stef and Nardie, brother and sister – well, who would have thought that we would be here today? My paranimphs Katrin and Maarten, for supporting me on this day.

Mikosch and Liene, my kids, for providing me with the necessary daily reminders that there are more important things to life than work. And, finally, Kata, for your love and support and, from time to time, just the right amount of pressure to keep me moving when the journey seemed to be all uphill.

And now we have arrived! Maybe not yet, as the poet says, ‘wise and full of experience’, but a bit wiser and more experienced nonetheless. It has been a pleasure and I am looking forward to our next trip.
Curriculum Vitae

Sander Hermsen, born in Gendringen, Netherlands, in 1973, is a behavioural scientist and designer. His research centers around two main lines:
1) The design and evaluation of products, services, and communication that support people in self-managing healthy, sustainable, and safe behaviour change.
2) Creating evidence-based and theory-driven methods and tools for non-behavioural scientists (such as designers and health professionals) to use theories from the behavioural sciences to inform their work.
Projects in both lines are typically multidisciplinary cooperations between behavioural scientists, scientists from other fields (e.g. medicine, physiotherapy, health and safety), civil servants, designers, and industry.

Education
2013 – 2019: PhD, Communication Sciences. VU University Amsterdam, Netherlands
2002 – 2007: BDes, Graphic Design, Artez School of the Arts, Arnhem, Netherlands
1992 – 1994: Communication Sciences, Radboud University Nijmegen, Netherlands

Employment
2011 – present: Utrecht University of Applied Sciences, Utrecht, Netherlands
Researcher Design and Behaviour Change at the Research Group Crossmedial Communication in the Public Domain
2018 (feb–aug): Radboud University Nijmegen, Nijmegen, Netherlands
Lecturer: Master's theses supervision for the behaviour change master track at the Behavioural Science Institute
2010–2013: Utrecht School of the Arts, Utrecht, Netherlands
Lecturer: Developing and teaching a course on Psychology for Interactive Designers and Game Designers
2007–2013: independent graphic designer, workshops on design and activism, support in action campaign design
Grants, exhibitions, publicity, honorary positions

November 2018: HU Utrecht personal post-doctoral grant
August 2018: SIA RAAK MKB grant (co-applicant) for Wat Beweegt Jou?!
February 2018: SIA KIEM Creative Industries grant for project In Control (applicant)
February 2018: SIA RAAK TopUp grant for SOLACE
November 2017: SIA KIEM Creative Industries grant for project Ontwerpen met Fundament (applicant)
October 2017: Behavioural Lenses toolkit exhibited in exposition De Nieuwe Stijl, Utrecht
July 2016: SIA RAAK TopUp grant for Touchpoints (applicant)
November 2014: Take it slow, Food Cognition and Behaviour grant awarded by NWO, Netherlands Organisation for Scientific Research (co-applicant)
September 2013: Doctoral grant, the effect of feedback through digital technology on habit disruption and change, Utrecht University of Applied Sciences (applicant)
April 2013: Touchpoints, two year SIA RAAK MKB research grant awarded by SIA (co-applicant)
September 2010: work shown at Neville Brody’s Antidesign festival, London
June 2009: external expert (formerly Rijksgecommiteerde) at the final exam of the Graphic design department, Artes Academie voor Beeldende Kunsten, Arnhem
November 2007: starter stipend BKVB fund (applicant)
September 2007: participant in the exhibition Arnhemse Nieuwe, selection of the most talented graduates of the design departments of the Arnhem college of art