Walking in the Wild – Using an Always-On Smartphone Application to Increase Physical Activity

Tim Harries[1], Parisa Eslambolchilar[2], Chris Stride[3], Ruth Rettie[1], and Simon Walton[4]

[1] Business School, Kingston University, Kingston Upon Thames, UK
[2] Computer Science Department, Swansea University, UK
[3] Institute of Work Psychology, Management School, The University of Sheffield, UK
[4] Oxford e-Research Centre, University of Oxford, UK
{T.Harries,R.Rettie}@kingston.ac.uk, csparisa@swansea.ac.uk, c.b.stride@sheffield.ac.uk, simon.walton@oerc.ox.ac.uk

Abstract. This multidisciplinary paper reports on a large-scale field trial, designed and implemented by a group of social scientists, computer scientists and statisticians, of a new smartphone-based app for the promotion of walking in everyday life. The app, bActive*,* is designed for a more diverse range of users than the typical active-lifestyle app, since it requires neither additional equipment nor a great deal of commitment to exercise. As a result, it can raise awareness of walking and promote walking amongst those with only a casual or hesitant engagement with the topic. The 6-week randomised controlled trial with 22-40 year-old male participants (N=152) indicates that bActive prompted users to increase the amount of walking they did by encouraging them to value and increase walking that is incidental to normal everyday activities. Longitudinal data analysis showed that use of the app increased walking by an average of 64% but did not find any evidence to suggest that the inclusion of comparative social feedback improves the impact of such apps on male participants.

Keywords: walking, feedback, norms, app, active-lifestyle, social sharing.

1 Introduction

Walking is generally considered enjoyable, relaxing [32], beneficial for general health [2] and helpful for the prevention of obesity and chronic disease [28]. It is "readily repeatable, self-reinforcing and habit-forming" [31] and is the most widely accessible type of exercise because it is inherently safe, requires no special skills, location or equipment and can easily be included in domestic and work routines [32]. However, most residents of advanced economies take far less than the daily total of 10,000 steps generally recommended for good health [5,10,20,29,33].

Unlike most previous apps aimed at increasing physical activity, the one described here, bActive, requires no special equipment and does not need to be activated by the

P. Kotzé et al. (Eds.): INTERACT 2013, Part IV, LNCS 8120, pp. 19–36, 2013.

user in order to track activity. This makes it easier to acquire (users only have to download the app and do not need to purchase any physical device), easier to use and more suitable for those with a passing or tentative commitment to becoming more active. The only demand made by bActive is that users engage with the feedback. This is encouraged firstly by their being able to do so at their convenience and, secondly, by the inclusion of trend data that is likely to engage their interest.

The bActive app collects data from the phone's built-in accelerometer and utilises an always-available display to deliver information to users about the number of steps they have taken. The presentation of this information was designed to increase walking activity in three principal ways. First, users are made more aware of the exercise involved in purposive 'walks' and that inherent in the walking involved in day-to-day activities such as shopping and work. Second, users are able to track their own activity over time. Third, they can be offered the chance to compare their activity levels with the average activity levels of others.

This study compared a control condition (no feedback) with the use of feedback limited to a user's own walking and the use of comparative data i.e. a *social norms approach* [11,18]. The social norms approach has been successfully used in fields as diverse as alcohol abuse, sexual behaviour, the payment of tax debts and domestic electricity consumption [9] but has not previously been applied to exercise or delivered by a smartphone app. The approach has two main elements. First, on the assumption that forces of conformity encourage people to emulate social norms, it provides individuals with information about the average behaviours of a group of salient others (the *descriptive norm*). Second, to avoid encouraging change in a negative direction (e.g. regression towards a lower activity level), some social norms practitioners provide users with moral approval for 'good' behaviour (the *injunctive norm*) [12,34].

A meta-analysis of the effectiveness of this approach in its main area of application, alcohol abuse by students, is presented by Bosari and Carey [6]. In this domain, the approach usually begins with survey research on respondents' own alcohol consumption and their assumptions about that of other students in the same university. Where this reveals a tendency to overestimate alcohol consumption, the discrepancy is conveyed to the student population of the university via poster campaigns. In another example, a randomized controlled trial of a US programme that posted reports containing social norms information with households' electricity bills found that this intervention reduced consumption by around 2% [1].

2 Related Work

Pedometers use piezoelectric accelerometers to count the number of steps walked and can be worn on the body or carried in a pocket. Pedometers have long been of interest for their ability to encourage more active lifestyles [3,16]. However, exercise promotion programmes have typically used pedometers alongside other resource-intensive activities such as classroom training and face-to-face sessions [7,23], so little empirical data exists regarding the use of pedometers in more natural contexts.

Pedometers exist either as devices dedicated to the measurement of exercise, or as embedded features in other equipment such as mobile phones. This difference has implications for accuracy and usability. Accuracy is highest amongst dedicated pedometer-devices, where it exceeds 96% at speeds of over 3 miles per hour (mph), dropping to 74%-91% at 2-3mph and 60%-71% at below 2mph [3]. However, users have to be committed enough to fitness and lifestyle-change to purchase the pedometer and remember to wear or carry it. This, and questions of fashion, design and convenience, can deter some people from using such devices [13]. In contrast, mobile-phone pedometer apps have the advantage of being embedded in equipment that people already own and keep on their person and their in-built display and communication capabilities allow users to share their feedback with others. Hence, smartphones are increasingly being used to address the problem of sedentary lifestyles [15].

2.1 Goal Setting

Goal-setting and performance feedback, which most generations of smartphones make convenient and easy,[1] are important influences on individual behavior [26] and are generally considered key features of technologies intended to encourage physical activity [13].

A number of factors influence the effectiveness of a goal. Firstly, it must be accepted by the individual and not in conflict with their other goals [27]. Secondly, increased difficulty is said generally to increase motivation [27], possibly because of the greater potential self-satisfaction that accrues to the user on achievement of the goal [36]. However, goals must not be considered unachievable [27] and though failure to attain a goal can increase motivation [8], it can also have the opposite effect [19]. Thirdly, the discounting of future benefits leads to the argument that short-term goals are more effective than longer-term ones [36].

One approach to determining physical activity goals is demonstrated by Chick-Clique [38] and UbiFit Garden [14,15], in which users set their own daily step-count goals. However, this approach runs the risk that inexpert users will set goals that are either too difficult or too easy and that do not, therefore, provide optimal motivation [33]. In contrast, in Fish'n'Steps [25] goals are set at a modest level by automatically using baseline step-counts as reference points and taking the findings of previous studies as a guide to what users could reasonably be expected to attain. Furthermore, Fish'n'Steps breaks longer-term goals down into proximal, daily, sub-goals. No research has been done to directly compare these two, contrasting, approaches.

2.2 Social Sharing

Mechanisms that facilitate social influence are also often considered essential for devices that encourage physical activity [13]. A number of smartphone apps include

---

1 See, for example, Nokia's Wellness Diary (`http://betalabs.nokia.com/apps/wellness-diary`)

this facility – e.g. [13,38,25]. In Fish'n'Steps, for example, each user is presented with a fish avatar whose growth, emotional state and behaviour reflect the number of steps the participant takes each day. The avatars of each group of users are displayed on a screen in an area shared by all its members (e.g. the social space in an office) and also on users' personal websites. A second example is MapMyWALK,[2] whose popularity is demonstrated by the fact that it had been downloaded from Android App Shop 250,000 times by July 2012. This app allows goals, routes, distances and walking speeds to be shared with friends and family members via email and social media.

Evidence on the effectiveness of these apps is mixed and small sample sizes cast doubt over its validity. For example, although Fish'n'Steps is reported to have caused some participants to increase their step-count, the unhealthy fish avatars that result from low activity levels caused others to drop out of the study. Similarly, although Consolvo et al [13] claim that social comparisons influenced their (all female) participants, the report of their three-week pilot of Houston with friendship groups of young females (N=13) reveals that the sharing of data did not have any significant impact on step counts. Finally, in a trial of Chick-Clique (N=7), group performance was reported by the participants (13-17 year-old girls) as the most "powerful method of changing behavior" [38; p1877], but the sample was too small to test this claim.

3 The bActive App

Although the design of the bActive app drew on the principles detailed in the previous section, it differs in three key ways from most of the apps described. Firstly, no assumptions are made about users' willingness to spend time and effort tracking their activity levels. Users simply have to download the app and carry their mobile with them in a trouser pocket. There is no requirement for additional equipment such as pedometers or foot pods, or for the data entry required for diarisation. While motivated users may be prepared to carry additional devices to measure specific physical activities, they are less attractive to those who are ambivalent about the benefits of measurement or their ability to become fitter and healthier [13], and hence are unlikely to promote ubiquitous use.

Secondly, unlike apps that activate only when users notify them at the start of an exercise event (e.g. MapMyWALK), bActive measures activity continually and without the need for any user action. Users are therefore rewarded with their activity data without having to make any initial effort or remember to switch the recording function on and off. This means, in addition, that rather than focusing uniquely on intentional exercise events such as hikes or walks, bActive also measures the exercise inherent in routine activities such as shopping, walking children to school or moving around at work. As a result, the emphasis is on the adoption of a healthy lifestyle, rather than on participation in walking as sport or recreation. This too is one of the

---

[2] `http://www.mapmywalk.com/`

reasons for not using GPS in bActive, for much incidental walking (e.g. shopping- and work-related walking) occurs in geographically confined spaces and is therefore less amenable to measurement by GPS.

The third difference concerns goal-setting. In bActive, formal goal-setting, training and coaching elements are replaced by self-generated, informal targets that result from a user's engagement with the feedback information. As a result, rather than feeling that they are engaging in a formalized exercise program, users are allowed to respond to this information in whatever way they wish. As argued by Thaler and Sunstein [37], behavioral feedback forms part of the choice architecture that nudges behavior. In this case, the bActive feedback nudges users to walk more. The only action required of them is that they occasionally bring the app to the foreground by clicking on the bActive icon, and this is subtly prompted by the presence of the bActive icon on the phone screen (see Figure 1).

To engage people who are initially less committed to increasing their activity, it is particularly important that bActive is seen as interesting and fun to use. Learning from Fish'n'Steps and UbiFit Garden, it uses non-literal, light-hearted visual representations of behaviour. It also provides trending information (as in Fish'n'Steps, UbiFit Garden, Into and Houston); gives positive reinforcement (learning from UbiFit Garden and Houston and from the problems experienced by Fish'n'Steps), and, like Houston, Chick Clique, UbiFit Garden and Into, provides opportunities for users to reflect on their own activity. Finally, like the social gaming and social data sharing features of Fish'n'Steps, Houston, Chick-Clique and Into, the social norms information within bActive is designed not only to prompt increased walking, but also to encourage engagement with the feedback.

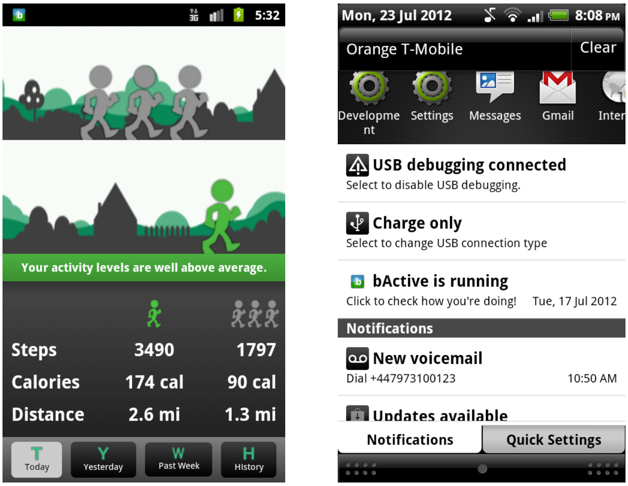

Fig. 1. Left: bActive's *Today* screen, as seen by those in the social feedback condition. (Note: the bActive icon is visible in the top left corner). Right: an entry on the Android notification bar indicates that the app is running in the background.

3.1 Design Philosophy

To test the efficacy of individual and social norms feedback, three different versions of the app were created: a 'null' version for participants in the control condition, which provided no user interface and gave no access to feedback; a 'partial' version for those in the individual condition, which displayed individual data only; and a 'full' versions for those in the social norms condition, which displayed both individual data and group averages. This section focuses on the version provided to those in the latter condition, for this was the most complex in terms of design.

Developed for Android 2.3, bActive incorporates automatic step counting alongside on-demand real-time and historic feedback of the number of steps taken by the user and the average of a group of other users. It also logs the frequency with which users open the app and the length of time the app is open on the display. The app is intended to be used on-the-go, so data clarity is emphasized and users are not asked to perform any retrieving or filtering tasks. To verify ease of use, the app was piloted twice and testing performed in a variety of outdoor conditions.

The app features four views of the data: *Today*, *Yesterday*, *Past Wee*k and *History* (Figure 1). The *Today* screen shows progress for the current day in terms of *steps*, *distance* and *calories* (calculated for the target demographic of young men) [17,39]. Values for the group average are displayed alongside those for the individual user. To facilitate rapid review of progress, an animated avatar representing the user (the green walking figure) is shown either behind, in front of or alongside an animated group of avatars (the grey walking figures) that represent the average activity for those in the comparison group. A banner just below the avatars displays a feedback message that varies according to how the individual's performance compares to that of the group. For those above average, it toggles between a descriptive norms message (e.g. "Your activity levels are above average") and an explicitly evaluative injunctive norms message (e.g. "☺☺ Well done, keep it up!"). If the individual's activity is below average, the banner displays a single descriptive norms message (i.e. "Your activity levels are below average"). The *Yesterday* screen is identical to the *Today* screen, but gives the previous day's results. The *Past Week* screen (Figure 2) displays a line graph depicting activity levels for the previous seven days, including averages for the group and for the most active 20%. Identical to the *Past Week* screen in format, the *History* screen allows users to swipe back and forth between different weeks.

3.1.1 Client/Server Architecture

The app uses the handset's mobile data connection to send users' activity levels and app usage to a central MySQL, and retrieves the group's activity average from that database. Step data is sent from the phone by a background Android service that transmits step data every two hours. This enables data to be updated asynchronously as users visit the screens, and ensures that they are shown up-to-date information. If lack of connectivity prevents data transmission, the app reduces the delay for activity transmission from two hours to one hour until a successful link is established. Data on use of the app is transmitted every four hours.

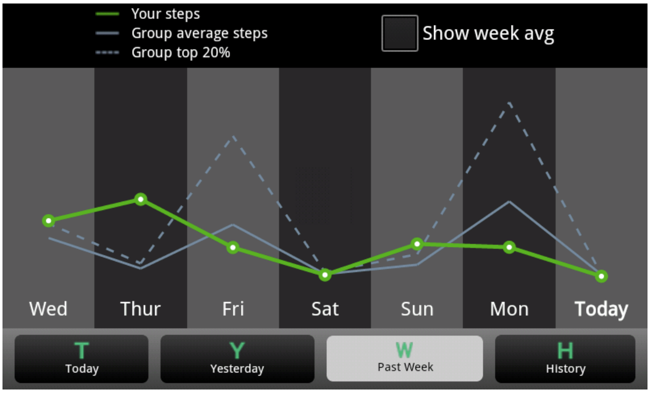

Fig. 2. The *Past Week* screen as seen by those using the full version of the app

A delay-based caching mechanism decides when particular elements of data on a user's device need to be updated and only requests fresh data when required. Meanwhile, the app provides the user with an indication of how much data it is using.

3.1.2 Activity Monitoring

The app's activity monitoring system is implemented as an Android service. Designed to be as autonomous as possible, it starts upon device boot and automatically recovers from unexpected crashes. On starting, it registers itself as a foreground service in order to prevent the Android process scheduler from hibernating the process when there are non-severe memory requests from other apps. An entry on the Android notification bar to indicate to the user that the app is running (Figure 1).

When the service is running, each accelerometer reading arriving as an *(x, y, z)* array is treated as a vector of magnitude $\boldsymbol{m}$. Instead of computing the true vector magnitude, the app simply computes m^2 as $x^2 + y^2 + z^2 - G^2$, where $\boldsymbol{G}$ is Android's accelerometer constant for Earth's gravity. The step counting algorithm is based on that of Mladenov and Mock [30], which treats accelerometer values as graph $\boldsymbol{y}$-values over time.

3.1.3 Keep-Alive / Battery Conservation Strategy

To preserve battery life, the Android power manager normally puts the CPU to sleep when the device screen is turned off through display timeout. Events such as phone calls and activated alarm clocks can turn the CPU and/or screen back on.

To ensure that the device captures a user's activity levels throughout the day without adversely affecting battery life, bActive uses a combination of two Android concepts: WakeLocks[3] and Alarms. Due to inefficient development of WakeLocks since Android 2.2,[4] a power strategy was used that allowed the CPU to sleep partially

[3] A mechanism to prevent the device from entering a low-power state.

[4] `http://developer.android.com/reference/android/os/PowerManager.html`

during periods of inactivity rather than reading the accelerometer at all times. The Android Alarm Manager wakes the phone from a low-power state once every 30 seconds and takes around twenty measurements from the accelerometer at a low frequency. If any of the readings are interpreted as representing a step, the accelerometer frequency is set to high and the device continues monitoring user activity until one minute after the final step is detected. At this point, the WakeLock is released, allowing the device to sleep for a further 30 seconds. Although essential for power saving, this strategy has the disadvantage that short bursts of activity (e.g. walking around a kitchen while cooking) will sometimes be missed if they are dispersed across periods of inactivity. This strategy may not be required in future generations of Android OS if the current issue with the WakeLocks is addressed effectively.

4 The Trial

Funded by the Research Councils UK Digital Economy Programme as part of the CHARM project, the six-week trial of the bActive app was conducted between October and December 2011 with 152 participants from Bristol, UK. Following the principles of a randomised controlled trial, participants were randomly assigned to one of a *control condition* (no feedback and no access to the interactive elements of the app), an *individual condition* (feedback on their own walking only – see Figure 3) and a *social norms condition* (feedback that also included social norms data).

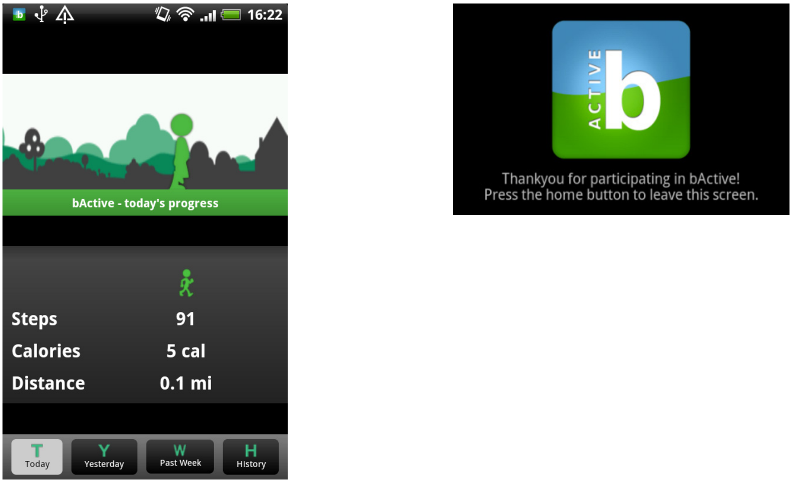

Fig. 3. The initial bActive screen as seen by those in the individual condition (Left) and control condition (Top)

On-street recruitment was conducted by a market research agency in twenty locations spread across Bristol. The incentive for participation was the study phone, the HTC Desire-S, which participants were able to keep. Recruits were told that the purpose of the study was to measure the amount of walking people did. Importantly, and unlike in most previous studies, participants were *not* asked to walk more. Each recruit had to agree to put his current SIM card into the phone, use it as his main mobile and carry it in his trouser pocket for the duration of the study. To ensure that users in the social norms condition were able to compare themselves to a group of broadly similar people, it was important to focus on a specific segment of users. Although the study could have been conducted with either men or women, it was decided to focus on men because of the need for the study phones to be carried in trouser pockets and the likelihood that women would find it more difficult to comply with this requirement. The age-range of 22-40 was chosen in order to avoid the health risks associated with the over-exertion of older users.

Of the 152 participants, 78% were in employment, 14% were students and 8% were unemployed. Of those in employment, 50% were in sedentary jobs (e.g. office workers), 44% were in moderately active jobs (e.g. teachers) and 6% were in very active occupations (e.g. postal workers). Meanwhile, 59% of the participants said that they regularly engaged in sport (about the same as the national average for this demographic [22]) and 63% owned some form of motorized vehicle.

The intention of the research team had been to collect a week of baseline walking data prior to the feedback phase of the trial, but a technical malfunction during this initial week rendered the data unusable, so no baseline comparison was possible.

During the 6-week feedback phase of the trial, participants were sent regular emails and SMSs reminding them to keep the phones in their pockets; with further prompts sent to those whose phones had not sent data for more than a day. In addition, those in the individual and social norms conditions were sent a weekly motivational text message (e.g. "Walking is one of the best activities for your health. How much are you doing? Check the app!")

During the trial, exercise and app use data was collected on a daily basis. Specific variables collected were the number of steps taken by each participant and, for those in the two experimental conditions that were provided with some kind of feedback, the number of times per day they activated the bActive app. In all, the resulting dataset consisted of up to 42 daily observations for each of the 152 respondents across the 6 week period of the trial (6,214 observations in total).

Prior to the trial and at its close, an online questionnaire collected demographic data and potentially confounding variables such as prior use of smartphones, patterns of physical activity, attitudes to physical activity and perceived impacts of the trial.

Subsequent to the trial, two waves of interviews were conducted with trial participants. Sampling criteria for wave-1 (N=7), conducted one to two months after the trial by a member of the research team, were feedback condition and self-reported changes in walking behaviour. For wave-2 (N=8), conducted ten months after the trial, sample selection focussed on those still using the app in April 2012 – four months after the end of the trial – and those with a below average total number of steps. Participants were approached by telephone and offered a £20 incentive.

Table 1. Features of the interview sample

Month of interview	Total	Feedback condition: C- Control I- Individual S- Social			Participant age		Living with chil-dren[1]	Self-reported change in walking[2]			Still using app April 2012	Self-reported comparison to average[3] A- above B- below C- varying		
		C	I	S	20s	30s		+	0	-		A	B	C
Jan 2012	7	2	2	3	5	2	3	4	2	1	n/a	1	1	1
Oct 2012	8	0	3	5	5	3	3	7	0	1	3	0	4	1

[1]The number of participants that had children living with them

[2]The number reporting that their walking had increased because of the study (+), stayed unchanged (0) or decreased (-)

[3]The number in the social norms condition reporting that during the study their step-count had usually been above average (A); below average (B), varied between above and below average (C)

The trial set out to test three hypotheses:

- H1 – those with access to feedback will have higher step-counts than those in the control condition
- H2 – those in the social norms condition will have higher step-counts than those in the individual feedback condition
- H3 – those in the social norms condition will use the app more often than those in the individual feedback condition

5 Analysis Method

Given the structure of the data, which had multiple daily observations nested within each participant, longitudinal multilevel modelling [35] was used to test the three hypotheses. As described below, the same 3-stage series of analyses was performed for H3 (where the outcome variable was frequency of app use per day) as was used for H1 and H2 (where the outcome variable was number of steps per day).

First, an unconditional model with no predictors was run in order to calculate the ICC(1) statistic (the percentage of variance in scores over time attributable to differences between participants) and the variance to be explained at the within- and between-participant levels. Second, a fixed growth model was fitted to the data in order to estimate the shape and direction of changes to the outcome over time. To do this, linear and quadratic effects of time (i.e. days since the start of the study) were added to the model, together with a dummy variable for day of the week (with Sunday set as the reference category).

The third stage tested for variability of change between participants by allowing the coefficients of the growth parameters (i.e. the linear and squared effects of time) to become random (i.e. to vary by participant). Potentially confounding participant-level control variables were then added: i.e. marital status; number of children under

17 in the household; employment status (30+ hours' employment per week; 8-29 hours' employment per week; carer/unemployed and in receipt of benefits; student, or self-employed); ownership and use of a motorized vehicle, motorcycle or bicycle (each coded separately), and previous ownership of a smartphone.

Finally, the effects of experimental condition and its interaction with the time-point were added to the model to see if they accounted for between-participant variation in the level of the outcome and its change over time. At each stage, model improvement was evaluated by testing the reduction in the model deviance and assessing the extra variance explained at both within- and between-participant levels.

There were two outcome variables. The first, Steps, was derived from the step-count, which was log-transformed to negate the impact of outliers (participants with extremely high step-counts). The distribution of the second outcome variable, App Use (the number of times per day users activated the bActive app), was severely positively skewed. For this reason, data for the participants who could access the app interface (the two experimental conditions; $N = 110$) were analyzed using a multilevel generalized linear model, treating the error distribution as Poisson and applying a logarithmic link function. Given the large sample size at the level of the time-point, study day ($N = 6214$ for Steps; $N = 4229$ for App Use), a significance level of $p < 0.0005$ was used for assessing the acceptance or rejection of null hypotheses at this level. For participant-level effects ($N = 152$ for Steps; $N = 102$ for App Use) the more typical $p < 0.05$ level of significance was applied. Where hypotheses were directional, one-tailed tests were used.

The interviews and focus groups were transcribed, coded using Atlas-ti and analysed using a combination of thematic and discourse analysis.

6 Results

6.1 Impacts of Feedback on Step-Counts

An assessment of the variation in *Steps* revealed a high level of clustering within participants, with an ICC(1) statistic of 0.33 indicating that a third of the total variation was due to between-participant differences. The introduction of linear and quadratic effects of time alongside dummy codes for day of the week explained a statistically significant but small 4% of within-participant variance and almost no between-participants variance. Tests of fixed effects coefficients indicated that of these three predictors (the linear effect of time, the quadratic (curvilinear) effect of time and day of the week), the third was the primary explanatory variable. Since quadratic change offered no improvement over a simple linear effect, it was dropped from the model.

There was evidence that the linear effect of time varied between individuals. When this random effect was added, along with the covariance between starting level and extent of linear change, the model deviance reduced significantly (Δ Deviance = 111 on 2df, $p < 0.0005$) and the unexplained within-participants variance was reduced by a further 4%. Of the demographic and control variables, only employment status and car ownership had a significant effect upon Steps, with full-time and part-time employees likely to have a higher step-count than other groups and car owners likely to have lower step-counts than non-car owners.

The tests for hypotheses H1 and H2 show that Experimental Condition (a dummy variable coded with control group as the reference category) had a statistically significant effect (Individual vs. Control: $B = 0.474$, $p < 0.05$; Social norms vs. Control: $B = 0.526$, $p < 0.05$), explaining a further 7.7% of the between-participants variance in the step-count. Compared to those in the control condition, the average expected step-count of those in the individual feedback condition was 59% higher and that for the social feedback condition was 69% higher (an average of 64% for the two experimental conditions). The null hypothesis for H1 was therefore rejected in favor of the finding that those receiving feedback had higher step-counts than those in the control condition. However, there was no significant difference in Steps between the two experimental conditions, so the null hypothesis for H2 could not be rejected.

There was no evidence that Experimental Condition had any effect on between-participants variation in the rate of change, over time, in the step-count. The interaction of experimental condition and time point was not statistically significant and this interaction reduced the model deviance by just 1 on 1df ($p > 0.05$), explaining only 0.6% of the variation in slopes.

Figure 4 shows the temporal variation in step-counts and illustrates the findings presented above.

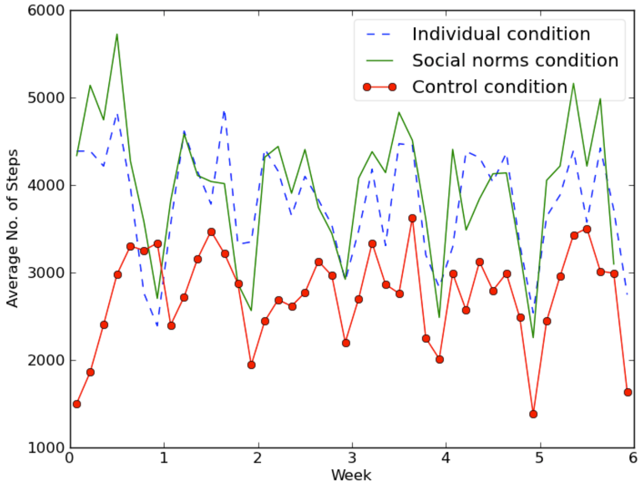

Fig. 4. Daily average step-counts for the three experimental conditions

6.2 Impacts of Type of Feedback on Engagement with the App

Over the course of the study, those in the two feedback conditions opened the app an average 3.9 times per day (median = 3.5, SD = 2.6), on each occasion keeping it open and visible on the screen for an average of 32.0 seconds (median = 33, SD = 9.0). Indeed, 87% of those in the individual condition and 89% of those in the social norms condition used it every day for half or more of the study days, and in the final week of the study participants from the two feedback conditions were still opening the app on average 2.3 times a day (SD = 1.9; median = 1.9). The survey evidence suggests that these figures reflect genuine enthusiasm for the app. Of those in the two feedback conditions 91% reported that the app was 'interesting', 67% that it was 'fun' and 73% that they would continue to use the app after the trial. Furthermore, only 19% reported losing interest in the app before the study end, only 15% of participants from the two feedback conditions reported that the step-count had not been "accurate enough for my needs" and only 11% that "lack of accuracy caused me to use the app less". The absence of any evidence for a non-zero correlation between perceived accuracy and either *Steps* or *App Use* indicates that problems with accuracy had little effect on the impacts of the app on behaviour.

As with Steps, variability in App Use was highly clustered within individuals.

As illustrated in Figure 5, the app was opened most often in the first few days of the study, with usage thereafter declining – first rapidly and then more gently (i.e. a curvilinear effect of study time-point was found to be statistically significant). This decrease over time varied between participants in both shape and rapidity, as evidenced by the fact that the addition of random effects of the time and time-squared terms (and the covariances between intercept and slope) increased the goodness of fit of the model (Δ Deviance = 1288 on 5df, $p < 0.0005$).

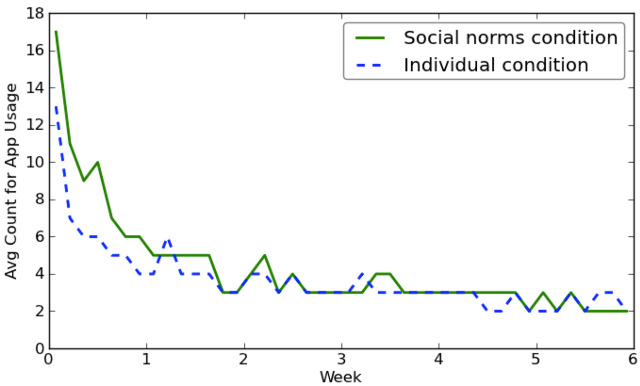

Fig. 5. The average daily frequency with which participants opened the bActive application on their phones

None of the participant-level control variables had a statistically significant impact on *App Use*. This is somewhat surprising, for men with children and those in employment (especially full-time employment) might be expected to have less time for such activities. The lack of any such effect suggests that interest in the feedback was sufficient for participants to use it in spite of other, conflicting, calls on their time and attention. The null hypothesis for H3 could not be rejected, for no difference was found between the two experimental conditions with regard to level or change of use.

6.3 Interview Findings

The interviews suggest that the increase in *Steps* by those in the intervention condition was the result of a number of features of the app design. Two of these were of particular importance: 1) the graphical display of walking patterns in the *Past Wee*k and *History* screens and 2) the always-on design, which meant that the app measured steps continually without having to be switched on. These features made users more aware of incidental walking. Interview respondents reported that before using the app they tended not to be aware of walking that was incidental to the achievement of other activities (e.g. the walking involved in shopping or work). The app made such walking more apparent, prompting one interviewee to comment, "...walking more than I thought, yeah. [...it is] surprising how much walking you do; just little bits here and there, walking around the workshop." This, in turn, showed users that they could become more active without having to engage in entirely new physical activities, and that all they needed to do was include more incidental walking in their existing activities.

The interview data further suggests that simply being seen to measure walking, the app encouraged users to view it as an activity in its own right. Where previously, walking had been "just something you do because it's a natural thing", it now became a measurable exercise that was subject to target-setting. Measurement also encouraged walking by helping users assess what was achievable. Before using the app, some users had relatively little understanding of distances, and the prospect of, for example, a two-mile walk might have been daunting ("Phew! Two miles, I'll hop on a bus" – David). By showing people how many miles they walked in the course of an ordinary day, the app made the concept of 'a mile' more familiar and thus made it more feasible for them to walk a number of miles ("Actually I walked two miles the other day and it seemed like nothing; I can walk that" – David). Finally, measurement highlights the difference between days with lots of incidental walking and days that are less active. This encourages users to walk more so that they are not obliged to see themselves as "lazy" or as "dossing".

The interviews do not fully explain the absence of an incremental impact for the social norms feedback. On the one hand, they suggest that those receiving social feedback became competitive and walked more when they thought they could "win" or "beat the average". On the other hand, there is some evidence that those who were below the average, either or a particular day or more generally, were less likely to walk more because they felt there was no possibility of 'winning'.

7 Discussion and Conclusions

This study indicates that always-on, accelerometer-based smartphone apps can increase walking amongst males by around 64%. This degree of behaviour change could have very real benefits for the general health of users and the prevalence of obesity and chronic disease in the population.

bActive's always-on feature allowed it to measure the walking inherent in practices not initially considered by users as 'walking' or 'exercise' and highlighted the periods in which they were relatively inactive. This had a transformative effect on some users, motivating them to avoid inactivity, making walking more of an activity in its own right, giving users the confidence that they could walk longer distances than they had previously realized and helping them to see that they could increase their activity levels simply by changing the way they conducted their usual activities.

Furthermore, there participants found the app engaging and enjoyable to use; with the trend data in the Last Week and History screens, in particular, holding their interest. Although use of the app fell away quickly from its initial high, the rate of decline slowed rapidly; participants were still accessing the app almost four times a day by the end of the study, and many expressed an interest in continuing to use it.

These results were achieved without the provision of any program of support or instruction, for apart from the feedback displays, the only input received by participants was the weekly motivational text message. This distinguishes the bActive evaluation from many previous studies, including many of those in the meta-analysis of pedometer interventions amongst the general population by Kang et al [23], the largest of which [21,40] included extensive additional programs of motivation or instruction.

The absence of any evidence in support of hypotheses H2 and H3 can be interpreted either as reflecting on the design of the social feedback or as evidence that 'social sharing' is not as important for the promotion of exercise as has previously been argued (e.g. [13]). The evidence in the literature on this issue is surprisingly weak, with assertions sometimes being made with little apparent empirical support. However, the lack of clear evidence for the effectiveness of social sharing does not allow its importance to be dismissed. Users were told that the comparison group comprised other males of about the same age who lived in Bristol, but it is possible that the salience and effectiveness of the comparisons would have been greater had users been given control over the types of people included. In addition, social sharing might have been approached in an entirely different way such as, for example, by allowing users to see the progress of other individual members of the comparison group.

A suggestion for future studies might be to increase the salience of the comparison group by defining it more tightly. Practitioners of the social norms approach generally argue that the most effective reference group comprises those whom participants consider most like themselves – e.g. [4,24]. This might, for example, mean separating those in physically active occupations from those doing more sedentary work.

A final consideration is the design of the feedback display. It is possible that the display of the social norms data on the same graph as the individual feedback detracted from the impact of the latter and that this, and not any lack of impact of the social data itself, weakened the impact of the full version of the app. Given that the individual feedback alone was associated with an increase in walking of 64%, it is clear that this should be an important element of any feedback strategy. However, when the social norm is much higher, even on just one day of the week, than a person's own highest step-count, the curve for the person's own steps becomes flattened and the all-important variations in his own performance are less obvious.

Minor issues with battery power and measurement accuracy notwithstanding, it is clear that when it comes to measuring and encouraging active lifestyles, the ubiquity of smartphones lends phone-based apps an important advantage over dedicated pedometer devices that require up-front commitment to increased exercise. Furthermore, the passive nature of data collection used by the bActive app encouraged use of the app and allowed users to collect data effortlessly and continually.

Those using the app recorded on average 64% more steps than those that did not. This highlights the power of the bActive approach and draws attention to the potential for the use of individual-level feedback that encourages people to reflect on the patterns in their own behaviour, identify opportunities for change and realise those opportunities. From the evidence in this study, it seems likely that an approach modelled on bActive could have a real positive impact on the health and fitness of the population.

Acknowledgements. The authors would like to thank the Research Councils UK Digital Economy Programme for supporting the bActive study and the wider CHARM project of which it was a part. Thanks, too, to Mubaloo, Bristol, for their constructive technical comments on the bActive app.

References

1. Allcott, H.: Social Norms and Energy Conservation. J. Public Econ. 95(9-10), 1082–1095 (2011)
2. Andersen, L.B.: Physical Activity and Health: Even Low Intensity Exercise Such as Walking is Associated with Better Health. Brit. Med. J. 334, 1173 (2007)
3. Baker, G., Mutrie, N., Lowry, R.: Using Pedometers as Motivational Tools: Are Goals Set in Steps More Effective Than Goals Set in Minutes for Increasing Walking? International Journal of Health Promotion and Education 46 (2008)
4. Berkowitz, A.D.: The Social Norms Approach: Theory, Research, and Annotated Bibliography (2004), `http://www.alanberkowitz.com`
5. Bohannon, R.W.: Number of Pedometer-Assessed Steps Taken Per Day by Adults: A Descriptive Meta-Analysis. Phys. Ther. 87(12), 1642–1650 (2007)
6. Bosari, B., Carey, K.B.: Descriptive and Injunctive Norms in College Drinking: A Meta-analytic Integration. J. Stud. Alcohol. 64(3), 331–341 (2003)
7. Bravata, D.M., Smith-Spangler, C., Sundaram, V., Gienger, A.L., Lin, N., Lewis, R., Stave, C.D., Olkin, I., Sirard, J.R.: Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review. J. Am. Med. Assoc. 298(19), 2296–2304 (2007)
8. Brusso, R.C., Orvis, K.A.: The Impeding Role of Initial Unrealistic Goal-Setting on Videogame-Based Training Performance: Identifying Underpinning Processes and a Solution. Computers in Human Behavior 29(4), 1686–1694 (2013)
9. Burchell, K., Rettie, R., Patel, K.: Marketing Social Norms: Social Marketing and the 'Social Norm Approach'. Journal of Consumer Behaviour 12, 1–9 (2013)
10. Choi, B.C.K., Pak, A.W.P., Choi, J.C.L., Choi, E.C.L.: Daily Step Goal of 10,000 Steps: A Literature Review. Clinical Investigative Medicine 30(3), E146–E151 (2007)
11. Cialdini, R.B., Goldstein, N.J.: Social Influence: Compliance and Conformity. Annu. Rev. Psychol. 55, 591–621 (2004)
12. Cialdini, R.B., Reno, R.R., Kallgren, C.A.: A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places. J. Pers. Soc. Psychol. 58(6), 1015–1026 (1990)

13. Consolvo, S., Everitt, K., Smith, I., Landay, J.A.: Design Requirements for Technologies that Encourage Physical Activity. In: CHI 2006, pp. 457–466. ACM (2006)
14. Consolvo, S., Klasnja, P., McDonald, D.W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., Landay, J.A.: Flowers or a Robot Army? Encouraging Awareness & Activity with Personal, Mobile Displays. In: UBICOMP 10, pp. 54–63. ACM (2008)
15. Consolvo, S., McDonald, D.W., Toscos, T., Chen, M.Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., Landay, J.A.: Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In: CHI 2008, pp. 1797–1806. ACM (2008)
16. Dewa, C.S., deRuiter, W., Chau, N., Karioja, K.: Walking for Wellness: Using Pedometers to Decrease Sedentary Behaviour and Promote Mental Health. The International Journal of Mental Health Promotion 11(2), 24–28 (2009)
17. Groves, A.: Using Pedometers as a Valid Method of Determining Physical Activity Intensity Level. MSc thesis, Brigham Young University, Dept. of Exercise Sciences (2008)
18. Harries, T., Rettie, R., Studley, M., Chambers, S., Burchell, K.: Is social norms marketing effective? A case study in domestic electricity consumption. European Journal of Marketing (in press)
19. Ilies, R., Judge, T.: Goal Regulation Across Time: The Effects of Feedback and Affect. J. Appl. Psychol. 90(3), 453–467 (2005)
20. Iwane, M., Arita, M., Tomimoto, S.: Walking 10,000 Steps/Day or More Reduces Blood Pressure and Sympathetic Nerve Activity in Mild Essential Hypertension. Hypertens. Res. 23, 573–580 (2000)
21. Jackson, E.M., Howton, A.: Increasing Walking in College Students Using a Pedometer Intervention: Differences According to Body Mass Index. J. Am. Coll. Health 57(2), 159–164 (2008)
22. Jones, H., Millward, P., Buraimo, B.: Analysis of the Taking Part Survey. University of Lancashire, Lancaster (2011)
23. Kang, M., Marshall, S.J., Barreira, T.V., Lee, J.-O.: Effect of Pedometer-Based Physical Activity Interventions: A Meta-Analysis. Res. Q. Exercise Sport 80(3), 648–655 (2009)
24. Lewis, M.A., Neighbors, C.: Gender-Specific Misperceptions of College Student Drinking Norms. Psychol. Addict. Behav. 18, 334–339 (2004)
25. Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., Strub, H.B.: Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 261–278. Springer, Heidelberg (2006)
26. Locke, E., Latham, G.: Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. Am. Psychol. 57(9), 705–717 (2002)
27. Locke, E., Latham, G.: New Directions in Goal-Setting Theory. Curr. Dir. Psychol. Sci. 15(5), 265–268 (2006)
28. MacLellan, G., Baillie, L., Granat, M.: The Application of a Physical Activity and Location Measurement System to Public Health Interventions to Promote Physical Activity. In: Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments, pp. 1–8. ACM, New York (2009)
29. McCormack, G., Milligan, R., Giles-Corti, B., Clarkson, J.P.: Physical Activity Levels of Western Australian Adults 2002: Results from the Adult Physical Activity Survey and Pedometer Study. Western Australian Government, Perth (2003)
30. Mladenov, M., Mock, M.: A Step Counter Service for Java-Enabled Devices Using a Built-in Accelerometer. In: Proceedings of the Fourth International Conference on Communication System Software and Middleware, pp. 1–5. ACM (2009)
31. Morris, J., Hardman, A.: Walking to health. Sports Med. 23(5), 306–332 (1997)

32. Pooley, C., Tight, M., Jones, T., Horton, D., Scheldeman, G., Jopson, A., Mullen, C., Chisholm, A., Strano, E., Constantine, S.: Understanding Walking and Cycling: Summary of Key Findings and Recommendations. University of Lancaster, Lancaster (2011)
33. Schneider, P.L., Bassett, D.R., Thompson, D.L.: Effects of 10,000 Steps per Day Goal in Overweight Adults. Am. J. Health Promot. 21(2), 85–89 (2006)
34. Schultz, P.W., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., Griskevicius, V.: The Constructive, Destructive, and Reconstructive Power of Social Norms. Psychol. Sci. 18(5), 429–434 (2007)
35. Singer, J.D., Willet, J.B.: Applied Longitudinal Data Analysis – modelling change and event occurrence. Oxford University Press (2003)
36. Steel, P., König, C.J.: Integrating Theories of Motivation. Acad. Manage. Rev. 31, 889–913 (2006)
37. Thaler, R., Sunstein, C.: Nudge: Improving Decisions about Health, Wealth, and Happiness. Yale University Press, New Haven (2008)
38. Toscos, T., Faber, A., An, S., Gandhi, M.P.: Chick Clique: Persuasive Technology to Motivate Teenage Girls to Exercise. Extended Abstracts on Human Factors in Computing Systems, pp. 1873–1878. ACM (2006)
39. Tudor-Locke, C., Ainsworth, B.E., Thompson, R.W., Matthews, C.E.: Comparison of Pedometer and Accelerometer Measures of Free-Living Physical Activity. Med. Sci. Sport. Exer. 34, 2045–2051 (2002)
40. Winnet, R.: Guide to Health: Nutrition and Physical Activity Outcomes of a Group-Randomized Trial of an Internet-Based Intervention in Churches. Ann. Beh. Med. 33(3), 251–261 (2007)