Happier People Live More Active Lives:
Using smartphones to link happiness and physical activity.

Neal Lathia[1], Gillian M. Sandstrom[2], Cecilia Mascolo[1], Peter J. Rentfrow[2]*

1. Computer Laboratory, University of Cambridge, Cambridge, United Kingdom
2. Department of Psychology, University of Cambridge, Cambridge, United Kingdom

* Corresponding author
Email: pjr39@cam.ac.uk (PJR)

Abstract

Physical activity, both exercise and non-exercise, has far-reaching benefits to physical health. Although exercise has also been linked to psychological health (e.g., happiness), little research has examined physical activity more broadly, taking into account non-exercise activity as well as exercise. We examined the relationship between physical activity (measured broadly) and happiness using a publicly available smartphone application. This app has collected self-reports of happiness and physical activity from over ten thousand participants, while passively gathering information about physical activity from the accelerometers on users' phones. The findings reveal that individuals who are more physically active are happier. Further, individuals are happier in the moments when they are more physically active. These results emerged when assessing activity subjectively, via self-report, or objectively, via participants' smartphone accelerometers. Overall, this research suggests that not only exercise but also non-exercise physical activity is related to happiness. This research further demonstrates how smartphones can be used to collect large-scale data to examine psychological, behavioral, and health-related phenomena as they naturally occur in everyday life.

Introduction

A sedentary lifestyle has been linked with many poor health outcomes. Time spent sitting is associated with increased risk of becoming obese, developing diabetes, cancer, and cardiovascular disease [1-3], and with increased risk of all-cause mortality [4]. Research suggests that -- beyond exercise -- small, cumulative, ‘non-exercise activity,’ such as standing and walking in the course of daily functioning, contributes to avoiding these negative outcomes and increasing general health [5]. A complementary question emerges: Are patterns of physical activity throughout the day also related to psychological health (e.g., happiness)?

To date, studies examining the relationship between happiness and physical activity have focused on exercise, finding mixed results. Some studies have found that happier people report exercising more [6, 7], while others have found no relationship between happiness and exercise [8, 9]. Much of this past research has relied solely on retrospective self-reports, on data collected at only one time period, and on small samples. Thus, no systematic research has examined links between happiness and behavioral markers of physical activity over time in a large, heterogeneous sample.

Every day, healthy individuals routinely engage in non-exercise physical activities. Because such activities are so common and, unlike exercise, do not require special planning or preparation, people may not be aware of the frequency or extent to which they stand, walk, or fidget throughout the day. As a result, assessing non-exercise activity using only self-reports risks generating coarse and inaccurate estimates. However, with the proliferation of smartphones that contain accelerometers -- sensors that detect the phone's movement -- it is possible to obtain objective estimates of physical activity throughout the day.

Ecological momentary assessment (EMA) is a powerful methodology for collecting in situ and in vivo momentary information about individuals’ psychological states, activities, and social contexts as they occur in everyday life [10]. Over time, such information provides rich (in the sense of quality) and dense (in the sense of quantity) information about individuals' naturally-occurring daily lives. Smartphones provide a useful platform for EMAs because people carry their phones with them most of the time, so they can complete momentary assessments anywhere and (almost) any time. Moreover, smartphones are equipped with an array of built-in sensors capable of collecting relevant data unobtrusively at very high resolution. For example, the accelerometer detects the phone's movement and can serve as a valid indicator of physical activity. Thus, with the combination of momentary self-report assessments and smartphone-sensor technology, smartphone-based EMAs offer a powerful platform for collecting data that, until recently, has been inaccessible to social scientists [11].

A number of studies have used EMAs to examine daily fluctuations in happiness. Results from these studies indicate, for example, that time spent with others is associated with increases in happiness [12]. However, all previous work in this area has relied on self-reports of both happiness and behavior. With respect to physical activity, most previous studies that have investigated a link with happiness have relied on self-reports of physical activity [8, 13, 14]. One study has measured physical activity objectively, using a physical accelerometer device attached to the hip [15]. However, the use of a physical device makes it challenging to recruit a large number of participants. In contrast, smartphone-based EMAs offer a valuable tool for investigating the link between happiness and physical activity in daily life from the large number of people who carry a smartphone, and with much more granularity than is possible with traditional, self-report methods.

The aim of the present investigation was to examine the link between total physical activity (including non-exercise physical activity) and happiness. We used longitudinal data from over 10,000 individuals who downloaded Emotion Sense, a publicly available mood-tracking smartphone application. Emotion Sense uses experience sampling (i.e., a form of EMA) to collect self-reports of happiness and physical activity, while at the same time passively collecting behavioral data from the accelerometer in the smartphone [16]. Several studies have demonstrated how the accelerometer can be used to detect the activities, postures, and movements of smartphone users [17, 18]. The integration of experience sampling and mobile sensing technology provides a powerful platform for collecting objective and longitudinal data at a large scale.

Materials and Methods

Participants

Participants were members of the general public who downloaded the freely-available Emotion Sense app, described below, from the Google Play store and installed it on their Android phone. The analyses reported herein include all users who provided data on all four happiness measures (described below) from February 2013, when the app was released, to June 2014, when we began the analyses. A total of 12,838 users completed at least one self-report survey; 10,889 of them provided demographic information. Of these, 43% were female, 54% were male, and 3% did not provide information about their gender. Users indicated their age by selecting a 10-year range. The mode was 25-34 years of age (N = 4,098; 38%), but large numbers of users were either just younger (15-24; N = 1,896; 17%) or just older (35-44; N = 3,426; 31%). A total of 85 users (1%) did not provide information about their age. More than half of the sample was White (N = 7,382; 68%). The next most represented ethnicity was Asian (N = 1,386; 13%), and 526 users (5%) did not provide ethnicity information.

Each analysis that follows refers to a different subset of users (e.g., those who provided information about some combination of happiness, self-reported physical activity and/or sensed physical activity). We thus provide further information in the Results section about the exact subset of users for each analysis.

The Emotion Sense Application

Emotion Sense is a smartphone application that was designed to study happiness and behavior. The app collects self-report data through surveys presented on the phone via experience sampling. By default, the app sends two notifications at random moments of the day between 8AM and 10PM, at least 120 minutes apart from one another. Clicking on a notification launches a momentary assessment, which includes measures of current affect (i.e., mood), and measures assessing a single aspect of current behavior or context (e.g., physical activity, location, social interactions). In addition to the notification-driven surveys, the app also allows for self-initiated surveys. These included longer measures of affect, and measures assessing multiple aspects of behavior and context. (Please see the Supporting Information for a complete list of self-report measures.)

As well as collecting self-report data, the app also uses open-sourced software libraries [19] to periodically collect behavioral data from sensors in the phone (e.g., accelerometer). The data collected through the app are stored on the device's file system and then uploaded to a server when the phone is connected to a Wi-Fi hotspot.

Emotion Sense was designed to be a tool to facilitate self-insight, providing feedback about how participants' affect relates to context and activity. When first installing the app, participants could only access feedback related to how their affect varies by time of day (e.g., morning, afternoon, evening, night). In an effort to maintain user engagement over a period of weeks, participants could receive additional feedback by ‘unlocking’ additional feedback screens, each of which had a particular theme (e.g., physical activity, location, social interactions) that determined which behavior and context questions were asked in the self-report surveys. Each successive feedback screen became available to be unlocked after a one week interval, and was unlocked by completing a measure of life satisfaction. The fourth screen that was available to be unlocked provided feedback related to physical activity. (Please see the Supporting Information for a complete list of self-report measures.)

This application was reviewed by and received approval from the ethics board in the Computer Laboratory at the University of Cambridge prior to release.

Measures

Happiness

Emotion Sense allows users to track and quantify their happiness in various ways. On each self-report survey, whether notification-driven or self-initiated, users indicate their current feelings by tapping on a two-dimensional affect grid (see Fig 1a), where the x-axis denotes valence, from negative to positive, and the y-axis denotes arousal, from sleepy to alert [20].

[insert Figure 1 here]

Fig 1. Self-Reported Mood in Emotion Sense. From left to right: (a) Measuring happiness in Emotion Sense using the affect grid: users select a point on a grid that quantifies valence (horizontally) and arousal (vertically). (b) Measuring happiness with PA/NA adjectives.

Users also rate their current positive affect (PA) and negative affect (NA) by indicating the extent to which various adjectives describe their current mood, using a 7-point sliding scale, with end-points at 1 (*Not at all*) and 7 (*Extremely*; see Fig. 1b). The mood adjectives, primarily drawn from the PANAS-X [21], correspond to each grid quadrant: high arousal/negative valence (angry, hostile, afraid, anxious, jittery), high arousal/positive valence (attentive, interested, alert, excited, enthusiastic), low arousal/positive valence (calm, relaxed, content), low arousal/negative valence (sad, lonely, depressed). On notification-driven surveys users rated two adjectives (see Supplementary Information for more details), whereas, on user-initiated surveys, they rated eight adjectives (two from each grid quadrant: anxious, angry; alert, enthusiastic; calm, relaxed; sad, lonely).

Finally, users also complete a broader measure of life satisfaction (SWLS, [22]) each time they unlock a stage of the app. Users indicate the extent to which they agree with each of 5 statements using a 7-point scale, with end-points at 1 (strongly disagree) and 7 (strongly agree). These were the only measures of happiness in the app.

For each user who responded to each of these measures (grid, PA, NA, SWLS) at least once, we created a happiness composite score. Following Diener and Seligman [9], we computed a z-score for (i.e., normalized) each of the four measures, which were moderately correlated with each other, $0.24 < |r'|s < 0.57$. Then we added together the z-scores for the user's average grid valence, PA, NA (reverse-scored), and SWLS. Users provided a median of 29 affect grid ratings (min = 2, max = 1369), 34 mood adjective ratings (min = 1, max = 1202), and 2 satisfaction with life ratings (min = 1, max = 23).

We validated this composite happiness score by examining its relationship with self-reports of users' average amount of laughing and crying, which could be reported in a separate part of the application (i.e., a between-subject analysis). Happiness correlated positively with laughing, $r(9,164) = .21$, $p < .001$, $d = .43$, and negatively with crying, $r(9,164) = -.18$, $p < .001$, $d = .38$.

Physical Activity

One way Emotion Sense assesses physical activity is through self-reports. Users indicate which activities they have been doing in the past 15 minutes (sitting, standing, walking, running, lying down, cycling or other; see Fig 2a). For each user we sum the number of active (walking, running, cycling) responses and divide this by the total number of responses (excluding 'other,' as it cannot be classified as either active or inactive), yielding an activity score that ranges from 0 to 1. For example, if a user reports walking, sitting, and standing in one 15-minute window, their activity score would be 0.33. Users reported sitting 8,127 times, standing 5,488 times, walking 5,399 times, running 1,076 times, lying down 5,708 times, cycling 533 times, and doing other activities 2,775 times. This is the only self-report measure of momentary physical activity in the app. (Please see the Supporting Information for a complete list of self-report measures.)

[insert Figure 2 here]

Fig 2. Physical Activity Data. From left to right: (a) How users self-report their recent physical activity, and (b) the magnitude of 30-second accelerometer samples collected on one device while performing each activity. The label (e.g., Walking: 2.713) contains the value of the feature computed from the given accelerometer sample; more physically demanding activities result in higher values.

Physical activity is also sensed via the phone's accelerometer. For 15 minutes before each survey notification and at regular intervals throughout the day, Emotion Sense measures the acceleration of the phone (see Supporting Materials for details about the factors affecting frequency of the accelerometer measurements). An accelerometer captures the acceleration the device is subject to, in m/s^2, in three dimensions (x, y, z). To score users' activity, we a) pre-processed the data, b) computed the axes' magnitude of acceleration ($x^2 + y^2 + z^2$; see Fig 2b), which is often used in activity detection [23], and c) quantified activity using the standard deviation of this signal (see

Supporting Material for details on this method and why we chose this particular measure).

Results

Validating accelerometers as a measure of physical activity

We measured the extent to which the smartphone accelerometer data aligned with self-reported physical activity. This analysis was run on the full set of self-reports of physical activity (i.e., each individual self-report for each user) that had corresponding accelerometer samples (i.e., samples at the same point in time as the self-reports; N = 23,419). A Pearson correlation found that the self-reported physical activity, referring to activity in the past 15 minutes, correlated with the activity score derived from the data sensed in the 15 minutes prior to the self-report, $r(23{,}417) = .37$, $p < .001$, $d = .80$. (We converted the correlation to a standardized mean difference with the formula d = 2r / sqrt(1-r2); seeSupplementary Information for analyses adjusting for possible lack of independence) This suggests that activity scores derived from smartphone accelerometers provided a reliable measure of physical activity, and therefore can be used as a coarse measure of physical activity in the absence of self-reports.

Are people who are more physically active also happier?

We normalized the happiness scores (see Measures section for details) across the set of users who had rated their happiness on all four of the measures contributing to the happiness composite score, and had provided measures of their average physical activity, either self-reported (N = 9,130), sensed (N = 10,371), or both (N = 8,737). We then correlated the normalized happiness scores with the average of the person's physical activity (i.e., a between-subjects analysis). The results indicated that self-reported physical activity was positively related to happiness, $r(9,128) = .08$, $p < .001$, $d = .16$, as was sensed physical activity (as measured by the accelerometer), $r(10,370) = .03$, $p = .002$, $d = .06$ (see Supplementary Information for results on individual happiness measures and results that control for personality). A regression predicting average happiness from both average self-reported and average sensed physical activity, entered as simultaneous predictors, found that both self-reported, $\boldsymbol{\beta} = .23$, $t(8,735) = 7.19$, $p < .001$, and sensed physical activity, $\boldsymbol{\beta} = .06$, $t(8,735) = 2.00$, $p = .05$, each independently predicted happiness. This supports the idea that objective measures can provide additional, unique insight into daily physical activity that goes beyond what we can learn from self-report measures alone. Taken together, these results suggest that happier people engage in slightly more physical activity (including non-exercise activity) than less happy people.

Diurnal patterns of activity

Next, we turned from examining a person's overall average physical activity to examining a person's average hourly behavior. For each user, we created two profiles of accelerometer data: one representing average weekdays and one representing average weekend days. Each profile is a vector with 24 entries, where each entry contains the average of activity scores captured in that hour of the day. We then used the k-means++ clustering algorithm [24, 25] to create three groups of users who exhibited similar diurnal profiles of activity (see the Supporting Materials for details on why we used $k = 3$).

Although we did not specify any criteria or thresholds for the groups, the k-means method identified groups of users who exhibited high, medium, and low levels of diurnal activity; each group's average activity is shown in Fig 3. Fig 4 visualizes the result by displaying the daily patterns of a random sample of 150 users drawn from each cluster. One-way ANOVA's, conducted separately on the weekday and weekend data, revealed that the clusters differed in happiness, both on weekdays, $F(2, 10{,}294) = 40.22$, $p < .001$, and on weekends, $F(2, 9{,}633) = 33.74$, $p < .001$ (results on individual happiness measures are reported in the Supplementary Information). Post-hoc Tukey's tests revealed that on weekdays, people in the high ($M = .25$, $SD = 2.86$) and medium ($M = .26$, $SD = 2.87$) physical activity clusters were happier than people in the low activity cluster ($M = -.27$, $SD = 3.12$), p's $< .001$, though the highly and moderately active groups did not differ from each other. On weekends, people in the high physical activity cluster were happier ($M = .57$, $SD = 2.81$) than people in the medium physical activity cluster ($M = .23$, $SD = 2.87$), $p = .002$, who were happier than people in the low physical activity cluster ($M = -.15$, $SD = 3.04$), $p < .001$.

[insert Figure 3 here]

Fig 3. Centroids for the clusters generated from (left) weekday and (right) weekend activity profiles.

[insert Figure 4 here]

Fig 4. A random sample of 150 users from each of the weekday and weekend clusters: users in the active clusters were, on aggregate, happier than those in the less active clusters.

Moreover, inspection of the clusters suggests that happy participants start their days earlier in the morning, end their days later in the evening, and display higher levels of physical activity throughout the day compared to less happy users. The levels of physical activity observed, however, were not intense or vigorous: Each cluster's average activity was smaller than the activity score that we manually collected while walking in a controlled setting.

Are people happier in the moments when they are more active?

Finally, instead of looking at average behavior, we took advantage of the repeated measurements we had for each person by running multilevel models, with momentary happiness and physical activity measurements nested within person (i.e., a within-subjects analysis). (NOTE: We couldn’t use the happiness composite as reported in the previous analyses because the life satisfaction measure was not reported on a momentary basis.) We did these analyses using the lmer package in R [26]. We ran separate models predicting each of the various happiness ratings (affect grid; high and low arousal, positive and negative emotions from the mood adjectives) from the z-scored self-reported physical activity, the z-scored sensed physical activity, or both. Given that multilevel models essentially require a minimum of 3 data points per person, these analyses were performed on the subset of users who provided at least 3 measures of both self-reported and sensed physical activity (N = 2,005).

Self-reported physical activity predicted more positive valence on the grid responses, more intense high arousal positive affect, and less intense low arousal negative affect. Similarly, when sensed physical activity was used as a predictor in a multilevel model, it predicted more positive valence on the grid responses, more intense high arousal positive affect, and less intense low arousal negative affect, (see Table 1; see Supplementary Information for results that control for personality).

Table 1. Multi-level modelling results predicting affect from physical activity.

	Entered Individually		Entered Simultaneously	
Measure	Self-reported physical activity	Sensed physical activity	Self-reported physical activity	Sensed physical activity
Grid valence	β = .04, *t* = 7.31***	β = .03, *t* = 6.06***	β = .03***, *t* = 5.29	β = .02***, *t* = 3.71
Positive Affect				
High Arousal	β = .09***, *t* = 13.17	β = .05***, *t* = 7.89	β = .08***, *t* = 10.92	β = .02***, *t* = 3.27
Low Arousal	β = -.02, *t* = -1.77	β = -.01, *t* = -.70	β = -.01, *t* = -1.54	β = -.003, *t* = -.31
Negative Affect				
High Arousal	β = -.002, *t* = -.27	β = -.002, *t* = -.22	β = -.002, *t* = -.17	β = -.002, *t* = -.23
Low Arousal	β = -.04***, *t* = -4.81	β = -.03***, *t* = -3.81	β = -.03***, *t* = -3.63	β = -.02*, *t* = -2.19

Degrees of freedom are 2,005 for grid valence, 1,996 for high arousal positive affect, 1,975 for low arousal positive affect, 1,958 for high arousal negative affect and 1,958 for low arousal negative affect.

* $p < .05$, *** $p < .001$

Further supporting the idea that both objective measures and self-report measures independently predict happiness, analyses that examined both self-reported and sensed measures simultaneously found that both self-reported and sensed physical activity predicted more positive valence on the grid responses, more intense high arousal positive affect, and less intense low arousal negative affect (see Table 1).

Discussion

Poor health has significant individual and societal costs. The current project showed that inactivity, which has been linked to poor physical health, is also linked to poor psychological health (i.e., lower happiness). Using a large-scale, public deployment of a mobile application that periodically assessed participants' happiness and passively measured physical activity, we discovered a modest but reliable association between happiness and physical activity. These findings have important implications for research on happiness, and also for behavioral science research methods.

It might be tempting to dismiss the results as spurious; when analyzing “big data”, such as the data used in this study, it is a common belief that *all* associations will be statistically significant. That is not our experience in working with the Emotion Sense data set. Also, it is important to point out that the opposite is also true: when analyzing small data, relationships can appear to be statistically significant when there is in fact no association. In fact, with the smaller standard errors that come with “big data”, we can be more confident that there is, in fact, a relationship between two variables, though that relationship may be modest. The association with happiness reported in the current paper is modest in size, but reliably manifests both for self-reported (i.e., subjective) physical activity and for objectively sensed physical activity. Obviously there are many factors that contribute to happiness. Given that positive social relationships may be the single most important factor contributing to happiness [27], we anticipate that the social interactions that underlie those relationships would have a strong influence on momentary happiness. Indeed people do report more positive affect when they are in social situations [28]. Future work that measures and statistically controls for this and other factors affecting happiness might reveal an even stronger relationship between happiness and physical activity. For now, given the size of the association, readers may wish to exercise caution when interpreting the importance of the relationship between physical activity and happiness.

The current findings extend previous research on the link between happiness and physical activity by demonstrating that regular physical activity -- including non-exercise physical activity like standing, walking, and fidgeting -- has a positive connection to psychological health. Experimental studies suggest that physical activity increases happiness [29, 30], but it is also possible that feeling happier leads to more activity. Although the design of the current study does not allow us to tease apart these competing hypotheses, the fact that our physical activity data were collected throughout the day suggests that happiness is linked to even non-exercise physical activity. Experimental investigations in which participants are instructed to increase or decrease their level of physical activity (both exercise and non-exercise) would shed light on the degree to which physical activity impacts happiness.

This study provides a compelling demonstration of how smartphones can collect large-scale data to examine psychological and behavioral phenomena as they occur in daily life [31, 32]. Indeed, mobile technology offered a methodology for investigating the link between physical activity and happiness that was more reliable and fine-grained than traditional methodologies. The results go beyond previous research using EMA for studying daily fluctuations in happiness by examining *objective* behavior. Nonetheless, it would be misleading to imply that smartphone-based EMAs can overcome all the limitations of traditional self-report and laboratory-based experiments. Indeed, there are broad and specific limitations associated with collecting data from smartphone sensors. Carefully controlled experiments are the cornerstone of science, and data collected in the wild are inherently messy and harder to interpret than data obtained from lab-based observations. Thus, smartphone-based studies allow for examining phenomena in the real world at the expense of losing control over all the events that naturally occur in life. More specific limitations are that even though smartphones are consistently proximal to their owners [33], there are likely to be instances where users are active without carrying their phones; the data in this study are likely to underestimate actual activity and to miss out altogether on rigorous exercise. Moreover, collecting continuous data from smartphones remains challenging, as this will quickly consume the device's battery. Smartphones are continuing to improve: more modern devices that have been released since our study include co-processors and dedicated physical activity tracking apps that are increasingly efficient at assessing diurnal physical activity. Future studies could make use of data from these apps (although some questions around accuracy remain [34]). Finally, it is also difficult to gauge the extent that receiving feedback from any app could alter both mood and behaviour [35]. Some of these limitations could be relieved by tracking daily activity with wearable sensors (see Kanning, Ebner-Priemer & Schlicht [36] for a review), but this approach remains, for now, expensive and impractical for assessing activity at the large scale necessary to detect the effects of small changes in behavior.

Conclusion

The frequency with which people physically move throughout the day, even if that movement is not rigorous exercise, is associated with both physical health and happiness. The current research reveals the important connection between physical and psychological processes, indicating that even slight changes in one has consequences for the other. The present findings validate the use of the smartphones to passively measure the daily activities of a diverse cohort of people, and pave the way for future research concerned with links between psychological and behavioral processes.

Acknowledgements

We thank L. Aknin, J. Crowcroft, S. Gosling, N. Lane, R. Lucas, M. Porter, P. Walsh, and A. Whillans for helpful feedback, and K. Rachuri for his contributions to the open-source software. This work was supported by the UK Engineering and Physical Sciences Research Council's UBhave project (Ubiquitous and Social Computing for Positive Behaviour Change, Ref. EP/I032673/1).

References

1. J. Levine, N. L. Eberhart, M. D. Jensen. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. Science. 1999; 283.212-214.
2. B. M. Lynch. Sedentary behavior and cancer: a systematic review of the literature and proposed biological mechanisms. Cancer Epidemiol Biomarkers & Prev. 2010; 19. 2691-2709.
3. E. G. Wilmot. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia. 2012; 55. 2895-2905.
4. V. der Ploeg, H. P. Chey, R. J. Korda, E. Banks, A. Bauman. Sitting time and all-cause mortality risk in 222,497 Australian adults. Arch Intern Med. 2012; 172. 494-500.
5. J. A. Levine. Non-exercise activity thermogenesis (NEAT). Best Practice and Research Clin Endocrinology & Metabolism. 2002; 16. 679-702.

6. J. Ruuskanen and I. Ruoppila. Physical activity and psychological well-being among people ages 65 to 84 years. Age and Ageing. 1995; 24. 292-296.
7. T. Stephens. Physical activity and mental health in the United States and Canada: evidence from four population surveys. Prev Med. 1988; 17. 35-47.
8. L. Catalino, B. Fredrickson. A Tuesday in the life of a flourisher: the role of positive emotional reactivity in optimal mental health. Emotion. 2011; 11. 938-950.
9. E. Diener, M. Seligman. Very happy people. Psychol Sci. 2002; 13. 81-84.
10. M. R. Mehl, T. S. Conner (Eds.) Handbook of research methods for studying daily life. New York: Guilford Press; 2012.
11. C. Wrzus, M. R. Mehl. Lab and/or field? Measuring personality processes and their social consequences. Eur J Pers. 2015; 29. 250-271.
12. W. Pavot, E. Diener, F. Fujita. Extraversion and happiness. Pers Individ Dif. 1990; 11. 1299-1306.
13. M. Wichers, F. Peeters, B. Rutten, N. Jacobs, C. Derom, E. Thiery, et. al. A time-lagged momentary assessment study on daily life physical activity and affect. Health Psychol. 2012; 31. 135.
14. Kanning, M., & Schlicht, W. (2010). Be active and become happy: an ecological momentary assessment of physical activity and mood. J Sport Exerc Psychol, 32(2), 253.
15. Kanning, M. Using objective, real-time measures to investigate the effect of actual physical activity on affective states in everyday life differentiating the

contexts of working and leisure time in a sample with students. Front Psychol. 2013; 3.

16. N. Lathia, K. Rachuri, C. Mascolo, P. Rentfrow. Contextual dissonance: design bias in sensor-based experience sampling methods. In ACM International Joint Conference on Pervasive and Ubiquitous Computing. Zurich, Switzerland. 2013.
17. L. Bao, S. Intille. Activity recognition from user-annotated acceleration data. Pervasive Comput. 2004; 3001. 1-17.
18. S. Hemminki, P. Nurmi, S. Tarkoma. Accelerometer-based transportation mode detection on smartphones. In ACM International Conference on Embedded Networked Sensor Systems. Rome, Italy. 2013.
19. N. Lathia, K. Rachuri, C. Mascolo, G. Roussos. Open source smartphone libraries for computational social science. Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication. 2013; 1. 911-920.
20. J. Russell, A. Weiss, G. Mendelsohn. Affect grid: A single-item scale of pleasure and arousal. J Pers Soc Psychol. 1989; 57. 493-502.
21. D. Watson, L. A. Clark. The PANAS-X: manual for the positive and negative affect schedule-expanded form. Ames: The University of Iowa. 1994.
22. E. Diener, R. Emmons, R. Larsen, S. Griffin. The satisfaction with life scale. J Pers Assess. 1985; 49. 71-75.
23. S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, M. Srivastava. Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks. 2010; 6.

24. D. Arthur, S. Vassilvitskii. k-means++: the advantages of careful seeding. Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics. 2007.
25. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et. al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12. 2825-2830.
26. Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-10
27. Lyubomirsky S., King L., Diener E. The benefits of frequent positive affect: does happiness lead to success? Psychol Bull. 2005;131. 803-855.
28. Pavot W, Diener E, Fujita F. Extraversion and happiness. Pers Individ Dif. 1990; 11.1299-306
29. S. Elavsky, E. McAuley, R. Motl, J. Konopack, D. Marquez, L. Hu, et. al. Physical activity enhances long-term quality of life in older adults: efficacy, esteem, and affective influences. Ann Behav Med. 2005; 30. 138-145.
30. C. Hansen, L. Stevens, J. Coast. Exercise duration and mood state: how much is enough to feel better? Health Psychol. 2001; 20. 267.
31. S. Gosling, W. Mason. Internet Research in Psychology. Ann Rev Psychol. 2015; 66. 877-902.
32. G. Miller. The smartphone psychology manifesto. Perspect Psychol Sci. 2012; 7. 221-237.
33. A. Dey, K. Wac, D. Ferreira, K. Tassini, J. Hong, J. Ramos. Getting closer: An empirical investigation of the proximity of user to their smart phones. In ACM

International Joint Conference on Pervasive and Ubiquitous Computing. Beijing, China. 2011.

34. M. J. Major, M. Alford. Validity of the iPhone M7 motion co-processor as a pedometer for able-bodied ambulation. J Sports Sci. 2016 May 30:1-5

35. S.D. Kauer, S.C. Reid, A. H. Crooke, A. Khor, S. J. Hearps, A.F Jorm, L. Sanci, G. Patton. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res. 2012; 14. e67.

36. M. Kanning, U. Ebner-Priemer, W. Schlicht. How to investigate within-subject associations between physical activity and momentary affective states in everyday life: a position statement based on a literature review. Front Psychol. 2013; 4. 1-16.

37. Kantar World Panel Smartphone OS Market Share. 2015;3; 16. Available: http://www.kantarworldpanel.com/smartphone-os-market-share/

38. M. Jokela, W. Bleidorn, M. Lamb, E. Gosling, P. Rentfrow. Geographically varying associations between personality and life satisfaction in the London metropolitan area. Proc Natl Acad Sci. 2015; 112. 725-730.

39. M. Kosinski, D. Stillwell, T. Grapple. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci. 2013; 110. 5802-5805.

Supporting Information

Materials and Methods

Participants

The Emotion Sense application was released to the public as a free download on the Google Play store in February 2013. A press release from Cambridge University in May 2013 and the subsequent coverage from a variety of newspapers and social media mentions helped to disseminate the app internationally.

Because Emotion Sense runs only on Android-powered phones, all participants were Android users. The demographic characteristics of the participants in our sample are comparable to the demographic characteristics of Android users [37]. Given that Android represents more than half of the smartphone markets in North America, Europe, and parts of Asia, there is little reason to suspect that our sample is not representative of the majority of smartphone users.

The Emotion Sense Application

Upon first opening the app, participants were presented with a consent form explaining the nature of the study and the data collected. The consent form assured participants that “any data this app collects will remain private and be securely stored'” and that “only members of the research team will be able to access your data.” Participants were also told they could request a copy of their data and have all or any part of their data deleted. Once participants provided consent, all the features of the app were activated and participants were prompted to read a short tutorial of how to use the app. When a user logged-out or uninstalled the app, all the features of the app were disabled.

In addition to the notification-driven and self-initiated surveys, the app also contained profile surveys. The app's “Profile” section displayed a set of surveys that participants could voluntarily complete. The profile surveys included demographic questions, as well as assessments of psychological constructs including personality, social connectedness, gratitude, and personal values, which users could complete once. None of the psychological assessments from the profile surveys were included in the present analyses.

As described in the main text, users could unlock a new feedback screen after each week of using the app. As expected, the number of users who unlocked successive feedback screens dropped. Although 10,818 users provided at least one self-report, 10,135 users provided self-reports at feedback screen 1 (time of day), 5,165 users provided self-reports at feedback screen 2 (location), 3,839 users provided self-reports at feedback screen 3 (SMS patterns), and 2,913 users provided self-reports at feedback screen 4 (physical activity). However, given that users also reported on physical activity during the self-initiated surveys regardless of what feedback screen they had unlocked, a total of 9,130 users provided self-report information about their physical activity.

Measures

Happiness

On notification-driven surveys, users rated their current affect on the grid, and then rated two mood adjectives. One of the mood adjectives was chosen from the same quadrant as the response on the grid. For example, if the user reported that their mood was negative in valence and high in arousal, then they were asked to rate one of the following adjectives, chosen at random: angry, hostile, afraid, anxious, or jittery. The second adjective was chosen at random from the list of adjectives for the remaining 3 grid quadrants.

Physical Activity

We pre-processed the accelerometer samples as follows. Some accelerometer samples returned little or no data. This could be due to (a) faulty sensors, or (b) some Android devices not generating accelerometer events when the screen is turned off. We first removed any samples that had no values in any or all of the three axes. We also removed samples that had fewer than 100 values per axis, which was deemed insufficient for a 10-second sample. Additionally, some accelerometer samples contained values that were impossibly high (e.g., acceleration values greater than 1,000); we removed samples containing feature values that were above 20, a value that is already substantially higher that what would be experienced by a typical smartphone.

The average time between users' first and last accelerometer sample is 51.9 days. The accelerometer sampling in Emotion Sense produces an average of approximately 20 minutes of sensor data per user, per day, uniformly at random. The app is set to collect sensor samples prior to creating a momentary survey notification and at regular 15-minute intervals throughout the day, but there are a variety of factors that further affect how much sensor data we captured from a user: users can manually increase/decrease the number of notifications they receive, some devices disable sensors when the phone is not in use, and some (although, we expect, few) users may manually disable the app's background sensing task.

We computed a large set of accelerometer features by extracting the mean, standard deviation and variance for each axis, and then using the same summary statistic to aggregate across axes, and then across a user's samples. We tested all of the extracted features by correlating them with self-reported physical activity. We chose to analyse the standard deviation of the magnitude since, compared to all the features we tested, it correlated most strongly with self-reported momentary physical activity.

Other self-report measures

Full details of all measures are available on request from the second author.

Notification-driven momentary assessments

Every momentary assessment included the affect grid, to assess mood, as well as questions that were tied to the stage the user was at. In Stage 1 (Time), in addition to the affect grid question, users also answered two questions about how their day was going. In Stage 2 (Location), users answered one question about where they were. In Stage 3 (SMS), users answered two questions about how active their social life had been. Stage 4 (Accelerometer) is the stage reported in the main text, in which users answered two questions about their physical activity. Questions in the remaining stages were only seen after completing Stage 4, so they would not have impacted the data reported in the current manuscript. The questions in these stages are as follows: stage 5 (Screen; two questions about device use), stage 6 (Microphone; one question about verbalizations, such as talking and laughing, and one question about noise in the environment), stage 7 (Calls; two questions about how active their social life had been), stage 8 (Personality; 5 questions about Openness, Conscientiousness, Extraversion. Agreeableness, Neuroticism), stage 9 (Sociability; 4 questions about recent social interations), stage 10 (Connectedness; one question) and stage 11 (Engagement; 5 questions about feelings of stress, challenge, creativity, interest, and enjoyment).

Self-initiated surveys

These surveys included questions about mood, as described in the happiness measures section of the main text. They also included all the questions from stages 1 through 7, as described above.

Profile surveys

These surveys were (generally) completed by users only once, and were not required. If a user had incomplete profile surveys, some text appeared on the app's home screen prompting them to complete the next available profile survey. In addition to the demographic measures and the life satisfaction measure, which are reported in the main text, users could complete measures of: personality (10 questions), gratitude (6 questions), health (8 questions), sociability (8 questions), job satisfaction (5 questions), life values (22 questions), and connectedness (10 questions).

Results

Happiness

We note that the users in this study were not especially happy; they scored lower on the SWLS (M = 19.40, SD = 6.54) than users who responded to an online survey hosted by the British Broadcasting Corporation (N = 588,014; M = 23.80, SD = 7.01 [38]). They also scored lower than users who completed psychological tests on the MyPersonality Facebook application (N = 101,068; M = 21.90, SD = 6.85; [39]).

There were minor differences in happiness between demographic groups in this study. For example, male users were happier than the female users, $F(1, 8{,}198) = 43.71$, $p < .001$. However, this difference was only apparent in the grid valence responses, F(1, 8,198) = 97.4, $p < .001$, and PA ratings $F(1, 8{,}198) = 131.7$, $p < .001$, but not in the NA ratings $F(1, 8{,}198) = 0.002$, $p > .25$ or SWLS $F(1, 8{,}198) = .81$, $p > .25$.

Further, for those age ranges that had more than 50 people (i.e., excluding people over the age of 70), a one-way ANOVA found that age predicted happiness, $F(6, 8{,}395) = 7.32$, $p < .001$. With the exception of the youngest users (less than 14 years old), mean happiness increased with age from age 15 through 64, though post-hoc Tukey comparisons revealed few significant differences between age groups.

Physical Activity

Overall, users did not report being highly active; the most frequent activities were sitting (43%), standing (21%), walking (21%), and lying down (14%), while both running and cycling each appear in less than 1% of reports, which broadly agrees with other studies on self-reported physical activity [31]. In 40% of the 555,436 activity responses received, users reported having been engaged in more than one activity. The most frequent pairs were combinations of the most frequent activities (sitting and standing, sitting and walking, standing and walking; for example, 43.2% of reports that included walking also included sitting).

Emotion Sense collected 109,306,542 accelerometer samples during the time under investigation. Much like the self-reports, an overwhelming majority of these samples indicate very low amounts of activity. Approximately 30% of samples have activity scores below 0.03, whereas when one of the authors manually collected walking data, the activity score was above 2.

Unlike the happiness scores, there were no age differences in self-reported physical activity, $F(4, 7361) = .64$, $p > .25$, but there were gender differences, $F(1, 7{,}320) = 18.72$, $p < .001$, with men reporting more physical activity ($M = .15$, $SD = .17$) than women ($M = .14$, $SD = .15$).

Validating accelerometers as a measure of physical activity

Although we have no reason to believe that the relationship between one self-report and one accelerometer sample is not independent from the relationship between another self-report and another accelerometer sample, even if both self-reports and both accelerometer samples come from the same user, for completeness we tested this relationship using multilevel modelling. We ran a model predicting self-reported physical activity from the sensed physical activity (level 1), grouped by user (level 2). The sensed physical activity is a significant predictor of self-reported physical activity, b = .16, $t(3{,}906) = 61.07$, $p < .001$.

The correlation between self-reported and sensed activity was similar in size for females, $r(10{,}394) = .36$, $p < .001$, $d = .77$, and males, $r(11{,}373) = .39$, $p < .001$, $d = .86$. These results held when tested with a multilevel model instead: b = .18, $t(1559) = 39.22$, $p < .001$ for females and b = .16, $t(1888) = 45.52$, $p < .001$ for males. Thus, the assumption that males may be more likely to carry their phones in their pockets while females may carry their phones in a bag does not seem to impact the validity of the sensor-detected physical activity data.

Are people who are more physically active also happier?

As we reported in the main text, using between-subjects analyses we found that self-reported physical activity was correlated with happiness (using the composite measure),

$r(9,128) = 0.08$, $p < .001$. If, instead of using the happiness composite, we look at the individual happiness measures, we see consistent patterns (see Table 2).

Table 2. Correlations between individual happiness measures and both self-reported and sensed physical activity.

	Grid valence	Positive affect	Negative affect	Satisfaction with life
Self-reported physical activity	$r=0.08$***, $d=0.17$	$r=0.06$***, $d=0.11$	$r=-0.05$***, $d=-0.10$	$r=0.06$***, $d=0.11$
Sensed physical activity	$r=0.03$***, $d=0.06$	$r=0.06$***, $d=0.11$	$r=0.02$**, $d=0.05$	$r=0.03$**, $d=0.06$

Degrees of freedom are 9,128 for self-reported physical activity, and 10,369 for sensed physical activity

** $p < .01$, *** $p < .001$

This effect did not seem to depend on the type of self-report; we find similar correlations when we look at only notification-driven or only self-initiated reports of physical activity (see Table 3).

Table 3. Correlations between self-reported physical activity and self-reported affect, comparing notification-driven vs. user-initiated self-reports.

	Grid valence	High arousal PA	Low arousal PA	High arousal NA	Low arousal NA
Notification-Driven	$r=0.09$***, $d=0.18$	$r=0.07$***, $d=0.14$	$r=0.05$**, $d=0.10$	$r=-0.05$**, $d=-0.10$	$r=-0.09$***, $d=-0.18$
Self-Initiated	$r=0.08$***, $d=0.15$	$r=0.10$***, $d=0.21$	$r=0.03$**, $d=0.06$	$r=-0.05$***, $d=-0.09$	$r=-0.06$***, $d=-0.12$

Degrees of freedom are 2,970 for notification-driven self-reports, and 8,986 for self-initiated self-reports

** $p < .01$, *** $p < .001$

Users could optionally provide self-reports of their personality. Of the 9,130 users who provided self-reports of physical activity, 890 also provided self-reports of personality. The relationship between average self-reported physical activity and average self-reported happiness (i.e., a between-subjects analysis), which was significant for the larger set of users, was marginal for the subset of users who provided personality data, $r(888) = .06$, $p = .07$, $d = .12$ (or, alternatively, for ease of comparison with the subsequent results, $\boldsymbol{\beta} = 1.34$, $t(888) = 1.82$, $p = .07$). Importantly, the size of this relationship was similar after controlling for personality; when average happiness was predicted from average self-reported activity and all of the big-five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism), activity remained a marginally significant predictor, $\boldsymbol{\beta} = 1.02$, $t(888) = 1.67$, $p = .10$. Each of the personality traits also predicted physical activity ($\boldsymbol{\beta} = .64$, $t(883) = 6.93$, $p < .001$ for Extraversion).

Of the 10,371 users who provided sensed physical activity data, 946 also provided self-reports of personality. The relationship between average sensed physical activity and average self-reported happiness (i.e., a between-subjects analysis), which was significant for the larger set of users, was not significant for the subset of users who provided personality data, $r(944) = .03$, $p = .38$, $d = .06$ (or, alternatively, for ease of comparison with the subsequent results, $\boldsymbol{\beta} = .79$, $t(944) = 0.88$, $p = .38$). Importantly, the size of this relationship was somewhat larger after controlling for personality; when average happiness was predicted from average sensed activity and all of the big-five personality traits, activity remained a non-significant predictor, $\boldsymbol{\beta} = .86$, $t(939) = 1.12$, $p = .27$. All of the personality traits, except for agreeableness, also predicted physical activity ($\boldsymbol{\beta} = .59$, $t(939) = 6.51$, $p < .001$ for Extraversion).

Diurnal patterns of activity

To analyse the effect of the k parameter on our clustering quality, we examined how the clusters produced from sensor data differed from one another with respect to self-reported physical activity, which was not taken into account when producing the clusters. We ran a series of one-way ANOVA's predicting self-reported physical activity from cluster membership. Post-hoc analyses using Tukey's test provided support for the choice of three clusters: When there are three clusters from the raw sensed data, they differ significantly from one another in self-reported momentary physical activity for both weekdays and weekends, p's $< .01$. In contrast, when there are four or more clusters, at least one pair of clusters do not differ.

We note that the vectors we clustered are potentially sparse, if no data was sensed/reported during those hours. We quantified the sparsity of each profile by counting the number of missing values per user. The profiles based on self-reported data have missing values for over 78.6% of users on weekdays, and over 83.9% of users on weekend days (spiking to 90% of users during night hours, from 1AM to 6AM). For those hours that do have data, the average number of reports is slightly higher for weekdays (M = 1.73, SD = 0.29) than weekend days (M = 1.29, SD = 0.11). On the other hand, profiles based on sensor data are less sparse: There are missing values for 42.5% of users on weekdays and 52.8% of users on weekend days. Bins that did have data contained a much higher average amount compared to self-reports: Weekdays had M = 50.05 (SD = 28.8) and weekend days had M = 24.82 (SD = 14.3) minutes of accelerometer data.

In the main text, we reported that when users were clustered based on sensed physical activity, they differed in happiness, as measured by the happiness composite. The clusters are also significantly different in happiness when we look at individual happiness measures (see Table 4).

Table 4. One-way ANOVA’s predicting individual happiness measures from cluster membership.

	Grid valence	Positive affect	Negative affect	Satisfaction with life
Weekdays	$F = 63.91$***	$F = 14.49$***	$F = 14.29$***	$F = 15.82$***
Week-ends	$F = 48.34$***	$F = 9.31$***	$F = 18.75$***	$F = 10.62$***

Degrees of freedom are 2 and 10,294 for weekdays, and 2 and 9,633 for week-ends.
*** p < .001

In addition to examining clusters based on sensed physical activity, we also ran a series of one-way ANOVA's predicting happiness, normalized across the set of users with happiness scores, from membership of clusters generated with self-reported profiles. We found a similar result to the clusters produced with sensor data. On weekdays the clusters differed in happiness, $F(2, 8{,}412) = 32.83$, $p < .001$. Post-hoc Tukey's tests showed that the least happy cluster ($M = -.10$, $SD = 3.05$) was less active than the other two clusters, p's < .001, which did not differ in happiness ($M = .50$, $SD = 3$, and $M = .50$, $SD = 2.86$). On weekends, the clusters also differed in happiness, $F(2, 5{,}613) = 14.76$, $p < .001$. Post-hoc Tukey's tests showed that the least happy cluster ($M = .04$, $SD = 2.99$) was less active than the most happy cluster ($M = .58$, $SD = 2.98$), $p < .001$, but neither differed in happiness from the third cluster ($M = .37$, $SD = 2.94$).

Are people happier in the moments when they are more active?

Of the 2,005 users who provided self-reports of physical activity, 380 also provided self-reports of personality. The multi-level modelling results remain similar when personality

	Entered Individually		Entered Simultaneously	
Measure	Self-reported physical activity	Sensed physical activity	Self-reported physical activity	Sensed physical activity
Grid valence	β = .05, *t* = 3.24**	β = .03, *t* = 2.10*	β = .04, *t* = 2.72**	β = .02, *t* = 1.17
Positive Affect				
High Arousal	β = .11, *t* = 7.17***	β = .06, *t* = 4.01***	β = .10, *t* = 6.28***	β = .03, *t* = 2.04*
Low Arousal	β = -.03, *t* = -1.86	β = -.02, *t* = -1.30	β = -.03, *t* = -1.53	β = -.01, *t* = -.92
Negative Affect				
High Arousal	β = .01, *t* = .59	β = -.004, *t* = -	β = .01, *t* = .70	β = -.007, *t* = -

traits are entered as simultaneous predictors (see Table 5). Personality traits also predict happiness; Neuroticism predicts all happiness measures, Extraversion predicts grid valence, high arousal positive affect and low arousal negative affect, and Conscientiousness predicts high arousal positive affect.

Table 5. Multi-level modelling results predicting affect from physical activity, controlling for personality.

		.30		.48
Low Arousal	β = -.04, *t* = -2.52*	β = -.02, *t* = -1.50	β = -.03, *t* = -2.19*	β = -.01, *t* = -0.78

Degrees of freedom are 380 for grid valence, 380 for high arousal positive affect, 378 for low arousal positive affect, 379 for high arousal negative affect and 377 for low arousal negative affect.

* $p < .05$, *** $p < .001$