Enhancing Understanding of Digital traces

Kristoffer Geyer

A thesis submitted for the degree of Doctor of Philosophy

University of Lancaster

Psychology Department

28/03/2021

Contents

LIST OF FIGURES

Chapter VII - Conclusion

List of Tables

ACKNOWLEDGEMENTS

From 2016 – 2021, I have been working on my thesis across an extremely chaotic period. Brexit, The Trump presidency, Covid-19 lockdowns & the great toilet paper shortage of 2020 were just some sources of disruption. All with vast implications for digital traces. Across the same period I immaturely inserted as much chaos into my own life as possible. I would like to take some time to thank those who provided some much-needed direction.

First of all, I would like to thank my thesis supervisor Dr David Ellis. Their own academic work is a source of inspiration. His critical but fair evaluation of my ideas on this thesis cannot be overstated. Additionally, the confidence that he placed in me really pushed me to develop as I would not have otherwise. Finally, I would also like to acknowledge their character for never saying “I told you so” after I did something he advised against, and it blew up in my face.

I would like to thank my family. My parent’s understanding of academia in their own fields has really provided a helpful context to understanding my own area. Their support during personal issues has been greatly appreciated as well. To my brother, who is starting work on their own PhD. I have enjoyed our debates on technology. We may always disagree on how unconstrained the license of technology developers should be, but these arguments have challenged me and questioned my ideas.

I would like to thank my partner Wenyu Huang. Her belief in me has provided me with boundless energy. Also, watching her dutifully & modestly carrying out her role as an infection control nurse during the pandemic has been inspiring. It is blatant to me that without your support and your example my thesis and my life would have been incomparably worse.

I couldn’t have gotten through this without the guys, thank you Danny, Alex & David. You’ve been my connection to reality across this time in academia. Anything really seems possible when you know that there are a few guys who always have your back.

Abstract

"How did Galileo demonstrate the veracity of the copernican view of the sun centred universe? Well, the main advances were incremental in their ability to refine glass into lenses, not very sexy, except he could use that to make his own telescopes and see the moons of Jupiter." - Dr Steven Hyman. (Cahalan, 2019, p 283)

Technological advances have repeatedly provided new tools for psychologists to conduct scientific inquiry. Theories regarding cognition, perception and behaviour have been more rigorously falsified or at least tested thanks to computers, electroencephalographs and associated big data sets. Smartphones are among the latest technologies being used for psychological research. Portable devices like these could provide unparalleled access into peoples' real-world behaviour via highly ecologically valid data. However, there are significant obstacles for psychologists to overcome before the potential of smartphones can be fully realised. Part one of this thesis documents multiple new methods that concern the development of smartphone apps for psychological research and provides guidance to ensure subsequent research is compliant with open science practices while maintaining participant privacy. In part two, developed apps are used in research designs that reveal inconsistencies between objective and self-report assessments of smartphone usage. Specifically, objective methods of measuring smartphone usage reduce the associations to almost zero between 'screen time' and health when compared with subjective estimates. This further demonstrates how the interdisciplinary application of smartphone technology can transform applied psychology or at least increase the methodological rigor.

CHAPTER I

INTRODUCTION

1.1 The role of technology in psychological research

A radical increase in speed of calculations, through the invention and advancement of computers, can revolutionise institutions across all areas of society. The earliest computer was employed during WWII, which contributed greatly to efforts in decrypting the communication of the axis powers (Hodges, 2012). The earliest IBM computer was highly disruptive for casinos in the 60s. Here, a mathematician used an early IBM computer to quickly perform years-worth of calculations in days to generate an optimum strategy for playing the card game blackjack/21 (Thorp, 1961). This forced casino owners to drastically reshape the 300-year-old game and a sizable financial return was netted by the inventor of the optimum-strategy (Thorp, 2017). The first adopters of machine learning in finance were Renaissance Technology (Zuckerman, 2019) who employed machine learning to identify hidden trends in markets and earn billions of dollars. Other areas have been slow to adopt but the impact has nonetheless been dramatic (certainly dramatic enough to be the subject of a movie; Miller, 2011). In sport, data mining allowed for the re-conceptualisation of baseball which ultimately, through appropriate changes in management and player behaviour, allowed an underfunded team to overcome all expectations and beat the richest teams in the league (Lewis, 2004). This approach would later be incorporated in how to manage a professional team in other sports and areas of life (Lewis, 2016). As I will demonstrate, the effect on psychological science specifically has been no less disruptive.

1.1.1 The history of computers in psychological research

The computer pervasively impacts all aspects of psychological research from: interactions between researchers, disseminating findings (e.g., via social media) and participant recruitment. However, the general psychological method has also been completely disrupted by computers. A variety of devices led to: adoption of more sophisticated methods of statistical analysis, improvements in experimental procedures, development of new types of experimentation and greater transparency in research. Here we shall explore the impact.

1.1.1.1 Analysis

Computers transformed the type of statistical arguments that psychologists and other researchers can make. Before computers were used in research, psychologists would use a hand-held calculator (or even an abacus) with pen and paper (Efron, 1979). Because the ability to perform calculations was limited by a researchers' mathematical ability and proficiency with calculators, the number of calculations that a statistical test could involve were limited. This limitation impacted the type of tests that psychologists could employ (parametric tests) and the assumptions that psychologists would have to hold about their data (the data would be normally distributed; although historic, political and cultural reasons also contributed to this assumption; Taleb, 2007). Even tests that were theoretically possible (e.g, Factor analysis) would require months of calculations conducted by hand (Dumont, 2010). Less rigorous tests were relied upon because they required less calculations to be performed. This includes "the multiple-group method" (Block, 1995, p 193) of Factor analysis. Additionally, simple mathematical errors are common and undermine the theories supported by the statistical test. Exactly this happened with a postulated structure of personality (Cattel, 1952; Dumont, 2010). After computers demonstrated their unparalleled ability to perform calculations, computationally expensive tests were conceived including bootstrapping, and error-rate estimation (Efron, 1979). Today, psychologists have been able to employ computationally heavy methods for modelling data such as deep neural networks (Ritter, Barrett, Santoro, & Botvinick, 2017).

Additionally, as computers have become more powerful they have better able to build more sophisticated models. While, it is a complex task to compare computational power, some useful metrics are available for understanding processing speed. These are floating point operations per second (FLOPS) and multiply-accumulate operations (MACCS; Hollemans, 2018). These metrics are both very useful to calculating capacity to develop machine learning models. The FLOPS are the number of floating point operation caried out per second, this is any arithmetic operation involving two floating point values. The MACCs are the accumulation of the values from two FLOP calculations. The conversion rate is the result of each FLOP must be summed and therefore the equation is as follows: MACC = 2FLOP – 1. If a model required multiplying 10 input values by 10 weights and summing the values then there would be 100 FLOPS and 49 MACC. IBM has been using FLOPS to calculate computing speed for a long time and this measure effectively captures capacity to run and test models (McCarthy, 2017). This metric is valuable as it is a deductive, gestalt summation how well all the components of the computer interact to perform tasks like develop machine learning algorithms, show animations, etc. As the duration required for a race car to finish a track captures the efficiency of the interplay between engine power, handling, breaking, etc., the FLOP calculations are a measure of computer performance from the interplay of the RAM, CPU, cooling system, memory handling etc.

The first super computer produced was the IBM 704, this had a FLOPS rate of 12,000 (McCarthy, 2017). The IBM 704 was only slightly less powerful than the Apollo Guidance Computer 14,245 (Averill, 2022). Computing speed has advanced and even common place pedestrian products, not designed for machine learning are comparably remarkable. A new games console was released, the Xbox X series (Microsoft, 2020), which can calculate a billion times more FLOPS per second than the IBM 704 (Versus, 2022a). The reported fastest computer is the Summit reported at peak to perform 200 petaFLOPs (Kumar, 2020) or 1667 times faster than the Xbox. But, how much does such changes impact psychology?

For a basic neural network there are two stages for the model that makes up the training (Brownlee, 2021): forward propagation, where the model makes a prediction from inputs; and back propagation, where the model identifies the errors and then makes corrections to the model. To take an example, a very minimalistic neural network (to keep numbers small) was designed to calculate if a face expressed one of eight emotions (Lisetti & Rumelhart, 1998). The model had one input layer, with 135 inputs. There was a hidden layer that had 40 nodes. Finally, there are an output layer with eight values being returned for each emotion. To calculate the MACC, we can simply multiply the inputs by the nodes in the hidden layers and then multiply that by the output nodes (135 * 40 * 8 = 43,200). Then to calculate the FLOPS we multiply by two and subtract 1 (43,200*2-1 = 86,399). This is the requirement for the forward propagation. Then the model calculates the size of the error and direction of error (the transfer derivative) this is 3 MACC per neuron in the hidden layer. Each neuron must be supplied with a transfer derivative to make a correction. The transfer derivative is multiplied against the summation of the predicted value against the intended output value. Then the neurons must be updated, this involves summing the learning rate, the past value of a given weight and associated input again this is 3 MACC. For the hidden layer each neuron must have the error calculated and the weight updated (40 * 3 + 40 * 3 = 240 MACC = 479 FLOP). Ultimately, for a model to make a prediction and make corrections on this the value (1 iteration) is 86,878 FLOP.

For a model to learn from 10,000 faces each make 8 separate expression and have 10,000 times being exposed to the dataset this would be 800 million times that the computer would have to front and back propagate (69.5 trillion FLOPs). Summit could achieve this in 3.6 milliseconds. This could be achieved by the Xbox in under 6 seconds. Whereas the if IBM

704 started developing the neural network as soon as it was invented (1954) we would expect the calculations to conclude around year 2137. It is very clear that the scope for model development with complex algorithms and linear models simply was not available with early computers.

The FLOP speed of supercomputers has increased exponentially at a rate of 1.398 every year(figure 1.1)

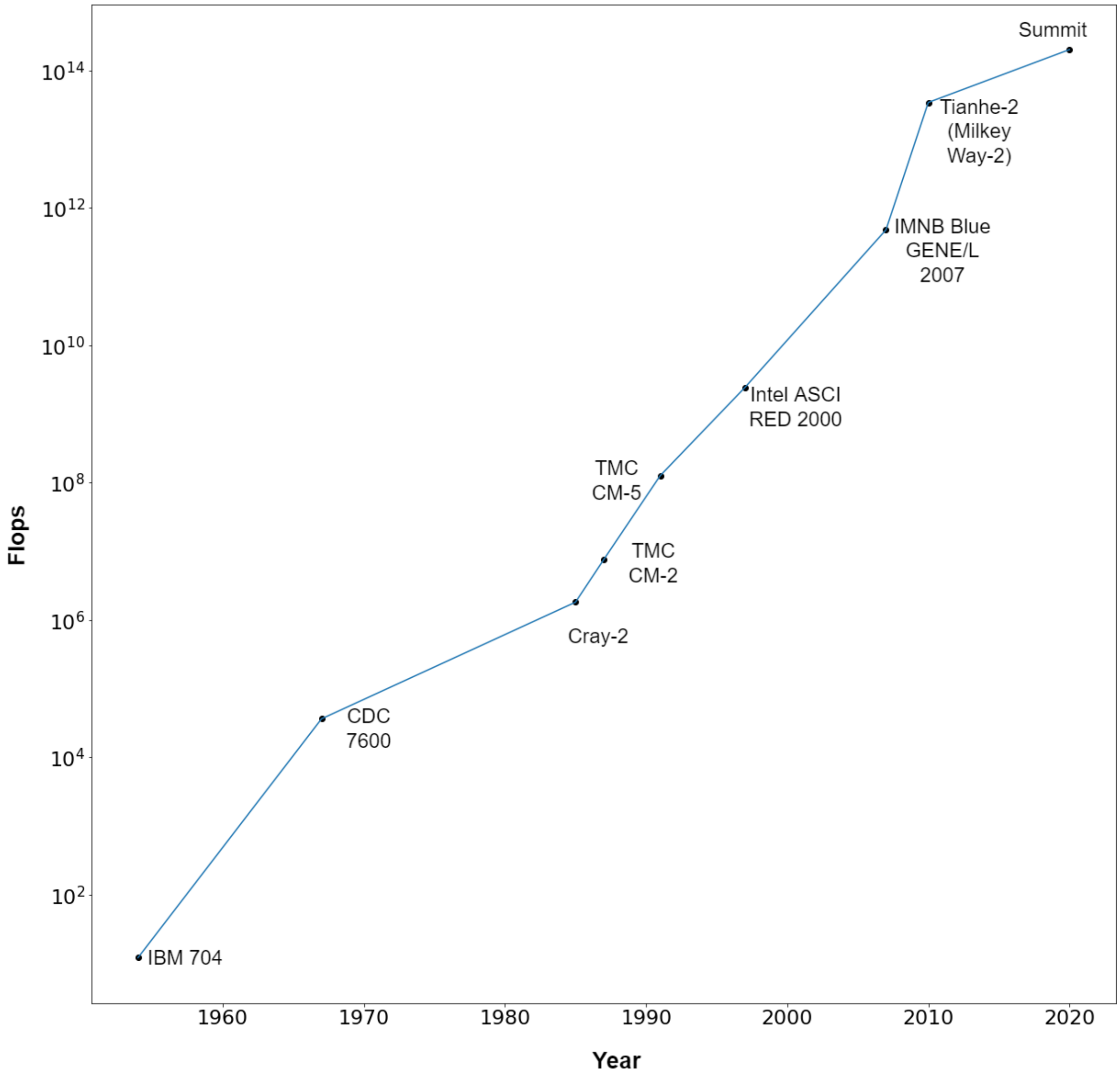


Figure 1.1 The exponential growth of FLOPS that a super computer can perform over the years (McCarthy, 2017; Kumar, 2020).

1.1.1.2 Improvements in experimental procedures

Computers have also allowed for enhanced control over stimulus presented to participants. The predecessor to computers in a laboratory setting was the tachistoscope which was a relatively basic tool (Buss, 1972). Previously, the tachistoscope could present two visual fields to a participant for an amount of time with millisecond accuracy (Taylor, & Maslin, 1970). However, the computer completely revolutionised the capability of researchers. Many visual fields could be presented, animations could be shown and paired with sound. The quality of the image was greatly improved. In the case of eye-tracking, researchers could programmatically ensure that the participant had viewed a particular object of a visual field for a required amount of time (Aaronson, Grupsmith, & Aaronson, 1976). Before, eyetracking would be carried out through using a video camera, but extensive time was required to analyse the data and errors would be introduced because of the tedium of the task. Computers could also facilitate complex interaction between participant and stimuli. Eventually, methods were invented to reduce the technical barrier involved with designing such a laboratory experiment on a computer (Norris, 1984; Bowler, 1987).

As time passed, computers would become exponentially more useful to psychologists. This was after the advancement of the internet, but this would not happen until the second generation of the internet. The first generation of the internet (Web 1.0) was generally when developers of website could produce content and visitors could see the content that was produced (Birnbaum, 2004). However, the relationship was one sided and the page would not allow for visitors interacting with the page in a meaningful fashion. Any interaction to a page would be completely forgotten when the page was refreshed and no other visitors to the webpage could detect that others were engaging with the page. In 1994, their programming language was updated (Hypertext markup language; HTML) which allowed people to view webpages, was upgraded to version 2 (Birnbaum, 2004). This allowed communication from the web page user to a designed server. Comment sections, blogs, social media and surveys were now a possibility. This innovation widely would be so dynamic as to revival the construction of the printing press (Silver, 2012). Quickly internet-based psychologists attempted to use this new tool for psychological research (Musch & Reips, 2000).

The internet was seen as a highly-efficient method for recruiting a more diverse sample of participants and not the most usual participant who "those who take lower division psychology courses" (Birnbaum, 2004; p. 820) later generalised to western, educated, industrialised, rich and democratic (WEIRD; Henrich, Heine, Norenzayan, 2010) individuals. Results gained from lab results were also replicated in a web setting (Birnbaum, 2004). Additionally, because the participant only needed to engage with a web-page and not interact with a researcher, an extensive amount of experiments could be run in parallel. Combined with an effective participant recruitment method, a decade's worth of lab work could be conducted in with minimal involvement from a researcher (Birnbaum, 2004).

Birnbaum (2004) did acknowledge dropouts in the experiment (as experiments were conducted in the time where internet connections were routinely interrupted due to a phone call) and participants may attempt to complete the experiment multiple times. Lip service was paid to the issue of variation in hardware and software, yet cognitive experiments often focus on reaction times and this could be compromised through variation in the computer specifications. But by far the most extensive limitation of this area of research was the technical skills required. Psychologists were expected to run their own website, this would not be easy. The experimenters would have to set up and house their own server, develop and publish the website, implement and manage their own data storage system and defend all of this against cyber security attacks. This probably requires competency with 5 separate programming languages: HTML, CSS, JavaScript, PHP & SQL; and a comprehensive understanding of server operations, cyber security, internet protocols and database management. If a psychologist mastered all of these technologies then a new threat was posed to their ability to do research. They may be headhunted as a web developer with a very substantial salary during the time of dot com bubble (Morris & Alam, 2012).

The first published experiment that collected data via the internet was published shortly after the emergence of the internet 2.0 (Welch & Krantz, 1996). This number snowballed until 2003 when there were 150 experiments listed (Birnbaum, 2004). However, this still remained small in comparison to other areas of research. These experiments did not regularly translate into publications. Only 22 papers (1.6%) of all papers (1401) in APA journals were collected via the internet (Skitka, & Sargis, 2006). Services like Mechanical Turk (mTurk; Amazon 2022) introduced in 2005 will have helped reduce or completely remove the barrier for entry to conducting experiments online. In 2011 61 papers were published using mechanical Turk but in 2015 this was 1200 (Bohannon, 2016). Psychologists represented the largest publisher of academic papers which utilised mTurk among all other disciplines. This shows that technical barrier to entry is an extensive obstacle for psychological research and this could well be an issue for adoption of psych apps as a method in psychological research.

1.1.1.3 Novel Experimental Procedures: Video games and Virtual Reality

Going further, video games have long been identified as a useful tool for psychological research (Jones, 1984). Early research into video games were identified generally for their potential as tools regarding performance and motivation to play the game (Arnold, 1976) related to how they recalled their success on the game (Isen, Shalker,Clark, & Karp, 1978). The first set of simulations of tasks would be programmed into a video game (Jones, Kennedy,

& Bittner, 1981) as a method of getting pilots experience without risking a plane crash and costing fuel. Then the concept of simulation as a method of practice spun out of this idea. Video games were also explored as a therapeutic intervention for patients with brain damage (Mickel, 1982). But there was a changing trend in the subject of interest for video games. From 1983 video games were primarily focused on video games as a learning aid. In 1994 video games were most commonly used as a tool for assessing bahviour, but into the 2000s psychologists were primarily studying the effect of video games on aggressive behaviour (Washburn, 2003).

There is still much interest in the utilisation of video games as an educational tool and much about research into gamification is underway. Additionally, the usefulness of educational video games has been shown as a useful tool for encouraging learning of languages such as Chinese (Rawendy, Ying, Arifin, & Rosalin, 2017), English vocabulary (Andreani & Ying, 2019), Spanish (Rachels & Rockinson-Szapkiw, 2018) and languages in general (Azzouz Boudadi, & Gutiérrez-Colón, 2020). Other skills include learning to program (Mi, Keung, Mei, Xiao, & Chan, 2018) and many subjects in education (Kusuma, Wigati, Utomo, & Suryapranata, 2018).

Research designs which were previously completely impossible were made possible by the increasing processing power of computers. Virtual Reality (VR) for example allows for complete control of what is presented to the visual and auditory fields (VR with haptic feedback has also recently been developed; BHaptics, 2021) of a participant while capturing the motions of the participant (Gaggioli 2001). The medium allows for psychological researchers to develop immersive virtual worlds for their participants (de Gelder, Kätsyri, & de Borst, 2018) where the laws of biomechanics (Linkenauger, Bülthoff, & Mohler, 2015) and physics (La Scaleia, Ceccarelli, Lacquaniti, & Zago, 2020) can be completely be rewritten. This singular ability to manipulate reality can permit participants to explore foreign virtual worlds and generally allow researchers to better understand how we navigate our daily lives. The method also offers precise monitoring of various behaviour: motion and eye-tracking (Yaremych, & Persky, 2019). Such areas of research that VR can explore include how participants would perceive their environment if their bodies were altered so that their ability to impact the environment was distorted (Linkenauger, Bülthoff, & Mohler, 2015)? For example, Linkenauger et al (2015) demonstrated that perception is impacted by the individuals' ability to interact with the environment through in/decreasing the size of the virtual arms of participants as a result their perception of their distance being narrowed or lengthened respectively. Other questions include: if a person was to live the life of another person would they be more ready to empathise with that person (van Loon, Bailenson, Zaki, Bostick, & Willer, 2018)? What impact does gravity have on visual perception (La Scaleia, Ceccarelli, Lacquaniti, & Zago, 2020)? Without VR and computers, we might struggle to ask such questions, let alone provide answers.

1.1.1.4 Reliability and Replication

While other aspects of psychological research may benefit from increasing in processing power of computers, for this area of research there are potentially new obstacles for replication. For context, the video game industry continually pushes what is possible for the newest game console. Indeed, each generation of console increases of the processing power of the last. This can be shown by considering the PlayStation from the classic PlayStation (Sony, 1994), PlayStation 2 (Sony, 2000), PlayStation 3 (Sony, 2006), PlayStation 4 (Sony, 2013) and PlayStation 5 (Sony, 2020).

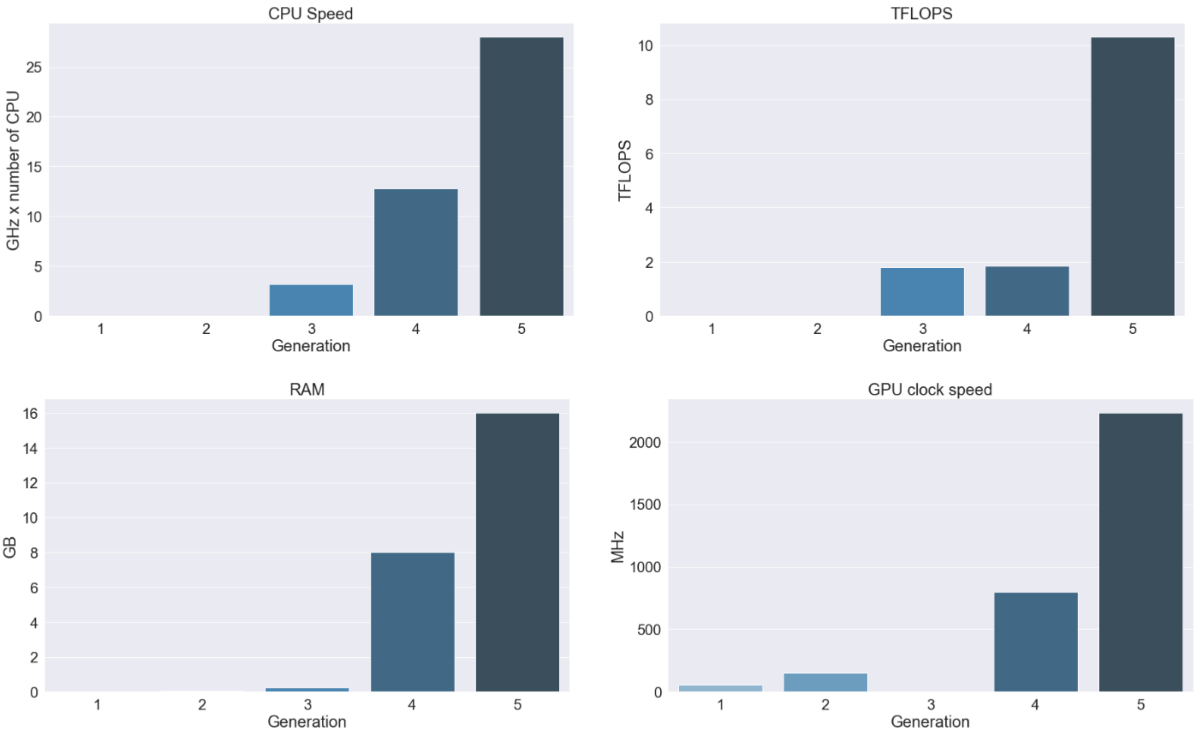


Figure 1.2 The difference of processing power for various PlayStations across metrics of CPU Speed, TFLOPS, RAM & GPU clock speed (Versus, 2022b). The TFLOPS of the PlayStation 1 are unknown as are the GPU speed of the PlayStation 3.

From the above it is clear that the processing power of the PlayStation 1 or PlayStation 2 hardly registers compared to the newest PlayStation 5 (What underpins this need for continual increasing capability of technology is addressed in a later section). This has real implications for the experience of the video game player. The games of even a decade years ago may not be sufficiently engaging for players to consider the experience immersive. The sophistication of image quality, game mechanics and virtual characters in the game are all directly linked to the performance power of the console. The power for each of these features all come from the same resources. An example of this was from the game Horizon Zero dawn (Guerilla Games, 2017), designers of the game had to cut features like the ability to have two players operate the game in order to make it visually appealing (O'Dwyer & Jayne, 2017). In summary, the more processing power relates to the more sophistication the medium across all aspects of the game. This has interesting implications for the replicability of experiments with VR and video games.

What is expected for a video game by an audience is continually changing and this may mean that the medium is difficult to research. Is there a problem for researchers looking to replicate findings with video games that there is a roughly 6 year pattern before a new console is brought out and the capability of the new generation of video games is revolutionized? Can this pattern limit the ability to replicate the findings from earlier studies? Certainly, studies like Arnold (1976) that report on asking undergraduates to use “complex video game” (Arnold 1976, p. 275) that graphically simulates a battle in Star Trek may not be consider complex in 2022. Therefore, the use of the same materials to investigate intrinsic motivation may result in a floor effect as participants generally are not intrinsically motivated to engage with the game.

The changing processing power required by video games can be shown with the contrast between final fantasy 7 (Kitase, 1997), the second most popular video game for the PlayStation 1 (Sony 1992) and then the remake 20 years later (Nomura, Hamaguchi, & Toriyama, 2020). The original was a super popular game and largely identified as a master piece of its time, the murder of the highly pixelated homunculus heroine is routinely rated as the most emotional scene by video game players (Russo, 2016; Dhar, 2020; Lykins, 2019). But the directors of the video game company must have decided that it was worth the time to make a whole new game from scratch with significantly improved graphics, more action in battle, and adding in audible dialog. Whereas simply converting the original game to being playable on modern consoles represents the smallest fraction of the costs of developing the remake. This suggests that modern young audience would no longer find the original engaging, the development of new content and not simply rebooting the original, supports this logic.

1.1.1.5 Transparency in Research

Blockchain technology is a promising technology for increasing transparency across research. Blockchain is generally a type of database which means that any changes to the underlying database is well documented (Nakamoto, 2008). Utilisation of cryptographic techniques ensures that every individual can only perform permitted functions. E.g. a person can transfer their digital assets but not others individuals' assets. Because of the clever utilisation of cryptographic concepts in this software the likelihood that an author of a change in the database is not actually that author is astronomically small: 1/2256 otherwise expressed as 1 in 115 with 75 trailing zeros. This database can also be distributed so that no centralised database exists but all individuals using the database can agree easily on which is the appropriate version of the database.

A smart contract (Zou, et al., 2019) is a program that runs within the blockchain database. It specifies conditions for what behaviour can happen within the database. Behaviour can include changes to the database, and values returned from the database. How this is different to any other program is that the program can be edited only once (when it is first uploaded to the blockchain) but anyone can read it. This is a digital version of terms and conditions of the program which are immutable.

Already, companies are springing up to offer blockchain based solutions for issues of transparency in academia (Mackey, Shah, Muyachi, Short, Clauson, 2019). Powering the blockchain technology is the bloxberg (Vengadasalam, Kleinfercher, & Lawton, 2020) becoming the standard blockchain for academic research. It has an interesting method of verifying that every change to the database is valid, which keeps the costs of changes to the database minimal. Essentially, as administrators' names are held in the public domain, if they attempt to manipulate the blockchain for private gain then their reputations will be tarnished

(Proof of authority consensus). This may be particularly effective consensus method in academia when the value of great academic work is often only fully understood after many years. However, request to designate authorship of work ideas occurs rapidly with this technology and therefore the value cannot be assessed before the transaction occurs; whereas with other blockchains, the value of a bitcoin or other cryptocurrency is immediately obvious. Typically, methods of developing consensus on cryptocurrencies are extremely resource intensive. The bitcoin uses proof of work consensus and relies on pure computing power for validation of transactions which is hugely resource expensive (Aste, 2016). Certain programs have been able to leverage the technology of bloxberg and build their own apps, which allows for collaboration between academics and version control of projects better documenting individual contributions of work (ArtiFacts, 2022). But generally, this only scratches the surface of issues with transparency in psychological research.

There are wide spread concerns in science with regards to reproducibility of science. Multiple landmark studies in psychology have been besmirched. Milgram (1978) omitted in his study of obedience that 72% of the participants were certain that they were not actually harming anyone (Hollander, & Turowetz, 2017). The study *On Being sane in insane places* (Rosenhan, 1973) was the subject of investigation (Cahalan, 2019) of an investigative journalist who raised serious doubts if the study was not an outright fabrication that extended out of only Rosenhan's experiences. Documents have emerged that suggest that Zimbardo and his research assistant had taken extensive steps to illicit violence and extreme behavior from their participants (Lapin, 2018; Blum, 2018). During the Stanford prison study (Zimbardo, Haney, Banks, & Jaffe, 1971) it has been alleged that Zimbardo pushed them to perform outrageous and horrific behavior indeed the designer of the experiment operated as a participant in the experiment. Others have been able to get away with careers of falsifying data (Marc Hauser - Carpenter, 2012; Diederik Stapel – Bhattacharjee, 2013). Largely, these are the scandalous cases relating to the lack of replicability and transparency of psychological research. Can blockchain systems exist to limit researchers' ability to game the academic system?

A concept of a blockchain based psych app is conceivable. Such a system may operate as follows: an app offers a platform for multiple different smart contracts that detail experiments. Such details include: experiment rational, what data is to be collected, the type of demographics of the participants, how many participants will data be collected from, and all the other usual considerations of the experiment. The researcher will have to submit a method for analysing the data prior to the experiment being launched. Once these elements have been submitted data from participants are collected as a part of the service. The data collected from participants is stored on the blockchain, but in an encrypted method (a side chain; Singh, et al 2020) to protect their data. Once the quota of data has been collected then the analysis is carried out and returned to the researcher. The analysis is also made publicly available but the raw data is not. The research is also provided a link to pass on to peer-reviewers where they can review how the app operates, the data collected and details of the underlying rational. If the underlying logic that runs all of these technologies is determined by smart contract then there is not a need to trust the developer of the software as the underlying code will be publicly accessible.

Generally, the exponential increase in computing power has been a liberating tool within psychological research, but has also brought some challenges. The devices have allowed for advancements in statistical arguments and from this theoretical freedom. The psychologists' laboratory experiment was initially enhanced by the involvement of computers. Then research was taken out of the laboratory and allowed recruiting people across the web. Advancements in VR allowed researchers to place participants into immersive virtual worlds and conduct research in such worlds, an impossibility without VR. Interestingly, computers historically have been unable to record real-world behaviour due to portability issues and a lack of sensors, which has created a disconnect between the digital world and the "real"/analog world. The static/virtual nature of computers is no longer a limitation however, due to the invention of the smartphone. Observational studies could completely be revolutionised if the data is generated by smartphone and not be labour intensive note taking of individuals' actions. The unavoidable interaction between observer and participant may be removed and thus demand characteristics minimized. One of the major questions that this thesis intends to answer is: Will insights delivered from smartphones come all at once or, as with previous evolutions, will there be a gradual growth as other areas of research come together?

1.1.2 Computational social science in psychological research

Computational social science involves using digital traces to study human behaviour (Conte, et al., 2012). Digital traces are generally any record that is stored electronically about persons' behavior either electronic or otherwise. The data employed by this method can be very diverse: phone records, bank records, public transport logs, health records, electoral records, social media data, google analytics, etc. This data is analysed by academics and data analysts within across multiple sectors and has also become fundamental to the modern economy. The data returned closely resembles what psychologists have been generating and analysing for the longest time: records of behaviour for a population or data describing the population. The methods allow for such large numbers of people to be sampled that inferences can be made about larger phenomena including spread of disease (Funk, Salathe, Jansen, 2010; Balcan, et al., 2009) and large-scale communication (Karsai, et al., 2011). This method does allow for analysing data about people's behaviour in the real world and perhaps has improved ecological validity compared to traditional experimentation. For example, researchers could capture the general sentiment of people tweeting through analysing a large number of twitter feeds and how the information spreads across networks by web scrapping a large body of tweets and analysing the results (Lerman, Ghosh, & Surachawala, 2012). This data was later used to generate effective predictions of general stock movements (Bollen, Mao, & Zeng, 2011). A Granger causality analysis was employed to infer a relationship between the general sentiment in the group and changes in the markets resulting in a suggesting that mood sentiment could predict with 87% accuracy if the market price would move up or down by end of trading on that day (Capelle-Blankcard, & Coulibaly, 2011). Note this data only tested the hypothesis from days on February 28th and December 19th 2008, these dates were during the great recession (Grusky, Western, & Wimer, 2011) and it was likely that the mood and stock market would be depressed. But computational social science replicated this data to detect an increase in liquidity in the housing market from 2008 to 2009 by measuring the frequency that people were using google to search for houses (Wu, & Brynjolfsson, 2015). Kosinski, Stillwell & Graepel (2013) demonstrated that Facebook likes could be used to accurately discriminate across certain groups based on sexual orientation in men (88% accuracy), racial background (95% accuracy) and political affiliation (85% accuracy). 61 million Facebook users (Bond et al., 2012) participants were either provided a list of their friends who had previously voted or not. If the list was provided then 0.39% more likely to vote than those who had not[1]. Evidence provided supported the idea of emotional contagion operates over social media like Twitter (Ferrara, & Yang, 2015). However, this next section will review the ongoing limitations that prevent the field from moving forward productively unless researchers take advantage of the humble smartphone, which can act as a hub for data transmission and collection.

1.2 Limitations of secondary data sets

Databases may not be easily repurposed by computational social scientists to adequately test new theories. For instance, psychologists have been interested in reviewing if

[1] It is unclear that such a small effect would have been identified as significant unless the study sample was so large.

more agreeable individuals have a greater number of friends. Agreeableness has been a long-standing trait in personality scales and captures the degree that others generally like an individual and therefore it would be reasonable to assume more likable people have a greater number of friends. Indeed, past studies show a link between number of friends and agreeableness in adolescents (Jensen-Campbell, et al., 2002). For more agreeable adults their friendship group consisted of non-biological relations (Wagner, Lüdtke, Roberts, & Trautwein, 2014). Indeed, a review of personality and peer relations found that agreeableness was associated with more friends, higher relationship quality and the agreeable individual was generally well liked (Harris, & Vazire, 2016).

Yet, when generating a model to predict agreeableness of Facebook (Facebook, 2021a) users, number of Facebook "friends" did not contribute to a model of agreeableness (Bachrach, Kosinski, Graepel, Kohli, & Stillwell, 2012). Indeed, while multiple features of a persons' Facebook profile were used to build models for predicting personality, only around one percent of the variance of agreeableness could be predicted. From this we may well question if Facebook "friends" are actual friends. Instead, Facebook friends would be better defined as a group of users who are designated with higher access (dependent on that person's privacy settings) to another user's data. Contradictory findings from virtual and non-virtual networks calls into question if findings from social media research can be generalised to non-virtual networks. Indeed, social media has been hailed as a platform for individuals to put themselves in the most favourable light and therefore may be difficult to discern what is contrived content from revealing of the individual (Stephens-Davidowitz, & Pabon, 2017). But researchers and analysts frequently to use such datasets not for their appropriateness to answer questions but their ease of access and use (Silver, 2011).

For instance, attempts have been made to try and analyse where people with types of personality typically spend their time by using the smartphone app FourSquare (Chorley, Whitaker, & Allen, 2015; Fouresquare, 2021; Noe, Whitaker, Chorley, & Pollet, 2016). This allowed for a list of locations that a person had documented accessing previously to be supplied to the researcher. The methodological problem involved the incentive structure involved of the app (West & Bergstrom, 2020). Participants would receive either financial or social rewards for accessing prestigious areas like fashionable restaurants. Participants would game the system and declare that they had accessed an area when they had simply walked past. Relying on social media apps for data sources ignores the interaction with the participant and social media site.

Previously, I have referenced a smartphones ability to act as a bridge between the virtual and non-virtual worlds. However, due to reasonable privacy concerns the user of the smartphone can act as a manager of what is recorded about themselves publicly. In the above examples, users of social media may well have inaccurately reported who their "friends" were (probably overreporting the size) and where they spend their time. There is much demand for providing the most flattering view of themselves on social media (Stephens-Davidowitz, & Pabon, 2017). Indeed industries are emerging which better empowers people to provide the most flattering view of themselves. B612 (Snow Inc. 2021a) a photo editor which can edit an individual's appearance (London, 2020) has been downloaded over 500,000,000 times on to android phones, the app is also used on apple (Snow Inc. 2021b) but the number of iPhone downloads is not published. Additionally, there can be real financial consequences for an unfavourable social media presence. Social media has been reported as a source of information used to determine credit score (Whateley, 2018). Individuals are also manipulating their social media to be perceived as desirable by potential employers (Roulin, & Levashina 2016).

1.3 Limitations when creating suitable measures

When a dataset offers significant detail, creating sensible measures remains challenging. For example, a person may enjoy watching a very specific set of films. As a result of this, making inferences about one person's experience and applying that to another can be notoriously difficult. Can the viewing history of individuals with similar tastes in cinema help make recommendations for other movies. For companies like Netflix (Netflix, 2021) it is a million-dollar question, literally! They hosted a competition for developers to build a better film-recommending algorithm and the prize money was $1 million (Netflix, 2009). The third-place algorithm pioneered an interesting approach by a solo developer who publicly shared his solution (Funk, 2006). Their solution was to take 8.5 billion rankings of every type of movie (supplied as part of the competition) on the database for the degree that it belonged to a genera. Then, they built an algorithm from the genre of the movie that the person typically watches and recommended other films in the same genre. This approach requires all data points (film) being understood as on a certain point on a number of different scales. To do such a task requires a huge amount of manual work from others to be successful (many reviews of each data point) or with Spotify (Spotify, 2021) a program is utilised to analyse every sub-aspect a song. Spotify for instance uses Echonest (2021) to extract over 500 features of a song. This can be added to a database to power a recommendation for what next to listen to. If a tool does not exist to make datapoint comparable then extrapolating data from them may be quite difficult. Again, when dealing with secondary data source, it may be very difficult to establish a scale which meaningfully allows all data to relate to all other points. For psychological science, we often don't have access to huge amounts of secondary data that can be repurposed in a way that suits our requirements. As before, when large volumes of data are available, they often require considerable unpacking before they can be useful and tables which join the data before it can become useful.

Smartphones however often allow a researcher to establish a meaningful relationship linking their datasets. A smartphone app used for psychological research for example, can retrieve extensive data about all installed apps including what permissions an app requests

(Android, 2021a). Permissions are the protected types of data that an app must request the user permit access to. An app is fairly academic if it does not request permission, it cannot access the internet, it cannot send texts or interact with another device, interact with any sensors (with the exceptions of some low-level sensors - e.g. accelerometer which in isolation reveals very little) or any data other that was not included in the app download, it is also unable to share any data with other apps. As an app without permission is extensively restricted, permissions are somewhat a declaration of what the app intends to achieve (Android, 2021b). As users can withhold some permission requests and still engage with an app, the capability of an app for one person may be different for another user. Messenger (Facebook, 2021b) may allow for calls to other Facebook users for some users but not others, therefore we can establish the affordances offered by a particular app for an individual user. This data can be used to quantify app functionality and group apps together. This ultimately quantifies what the app is capable of and the affordances that it offers to the user relative to other apps.

Comparatively, a social scientist may only be able to seek out public databases and search out attempts to group apps. Therefore, there may not be any datasets that provide effective context on the significance of the data and how different data points relate. For instance, attempting to understand what an app offers a participant may be unclear from the tags that are supplied on app stores. They may use an API (MightySignal, 2021) to get lists of how apps present themselves on App Stores but the public presentation of apps may have a tenuous link with what is actually offered by the app. The radically different apps: Tinder (Tinder, 2021), Subway (Subway, 2021), Tesla (Tesla, 2021), H & M (H&M, 2021) are all in the same category - Lifestyle. Therefore, such genera categorization of apps could not distinguish between someone looking to date, use a car or have a sandwich. Indeed, branding may change independent of changes in the affordances that an app offers.

This can also occur when making sense of smartphone sensor data. Computational social science may possess the transport logs of an individual. However, smartphone data would be able to better explain the context of each time a person used public transport. What localities were they accessing to get to public transport, was it part of a longer journey, do they typically use a car, is the area they access more lively/noisy than they usually access (established through a microphone), is it busier (through utilising Bluetooth) etc. (A demonstration of this is provided in the conclusion). Therefore, the ability to make sense of data goes far beyond what is typically possible with computational social science.

1.4 The potential contributions of smartphones to psychological research

1.4.1 Characterisation of a smartphone

The iPhone was perhaps the first smartphone as we currently recognise them. Phones which connected to the internet had existed before (indeed 1992 saw the first mobile phone which was capable of connecting to the internet and therefore a mobile phone with the "smart" component; Merchant, 2017), but they lacked key features which meant that the phone was not really a smartphone as we now understand it. The iPhone pushed the extent that phones were reprogrammable. Given that a physical keyboard was not required on a device meant that developers could utilise the entirety of the screen (this was spelled out in the iPhone release; Jobs, 2007). Everything on screen could be altered dependent on what the user's purpose was and the apps they downloaded. Third-party developers were invited to develop apps which could be used on the iPhone. This drastically lowered the bar for requirements in order to produce apps for phones. Prior to the iPhone, only organisations producing apps would be companies which built the phones. While these companies all seemed capable of building a monochrome snakes moving around the screen, they did not seem to want to deliver anything much more innovative and this really limited the extent to which the devices were actually reprogrammable. Now, content for smartphones would be produced by entities which were not well-established smartphone producers. All of this together decentralised the development of smartphone apps. Anyone could build an app, which anyone could use anywhere. Although, issues of infringements of competition law are currently being debated (BBC, 2021). Demonstrations of the issues regarding competition and gatekeeping of content are apparent as the video game Fortnite (Epic Games, 2021) was banned across smartphone platforms after trying to sell virtual currency independent of the smartphone vendors (Paul, & Sweney, 2020).

Another significant feature of the iPhone is the in-built sensors that can communicate real-world information about the status of the user and allowed for the blending of digital and real-world interaction through using sensors like GPS sensors, accelerometers & geomagnetic field sensor. Google Maps (Google, 2021a) demonstrated the new affordances were offered by the smartphones. The development of Google Maps was heavily influenced by a developer who previously pioneered methods of information integration and data visualisation for generals in the military (Merchant, 2017). As the smartphone could access data on network connections, WiFi or GPS; Google Maps could quickly get the user's location (Packages made this simple for developers, see chapter II). Therefore, a smartphone for the most part would immunise the user from getting lost. But the affordances that this offered quickly multiplied. Third-party apps could programmatically establish location meaning calculations could be carried out regarding the space between the devices of users, this allowed for connecting people in new ways. Early third-party apps which based on the location started to emerge early on with FourSquare (2021), Grindr (2021) later apps would also utilise the location sensing capabilities as central to their program: Pokemon go (2021), MapMyRun (mapMyfitness, 2021) & Detour (2021).

The smartphone generally is now quite a powerful device and could easily achieve historical computational achievements. The code for the apollo space guidance system is publicly available (Iravania, 2019) and easy enough to implement with an android. Similarly, the code to the generate enigma level encryptions and crack the code is available here (Pound, 2021). Similarly, the new Microsoft Lumina 950 XL (Versus, 2022c) a very phone with a very powerful processor for a smartphone. As we've established, the IBM 704 could build a neural network in 183 years, a new Xbox can do it under 6 seconds and the new Microsoft phone can do it in just over 3 minutes.

1.4.2 Smartphone's promise of access to real world data

Smartphone apps that are designed to conduct psychological research/interventions are called "psych apps" (Miller, 2012, p 221). They can perform ecological momentary assessment by providing questionnaires to the participant in a real-life setting. While this has its place in the methodological toolbox for psychologists, the full potential of smartphones is further realised when continual observations of a user's actions are monitored during the research without intervening in the participants life and collecting behavioural data in the background (Miller, 2012). Cattell (1958) argued that people always try to appear in the best light. Very few people would be comfortable allowing negative impressions of themselves being scientifically established (no one wants the conclusion of an empirical investigation to be that they are not a nice person!). Therefore, the best evidence for identifying aspects is where the participant has least control over how they appear to the experimenter. He suggested that questionnaires are simple to manipulate, as ticking one box is as easy as ticking another. Laboratory experiments were generally harder to manipulate as the dependent variable or independent variable were not generally easy to ascertain. Cattell considered observational studies the hardest to manipulate as behaviour had to be changed in the real world. Behaviour would have to be added to or subtracted from. A person attempting to appear dedicated to exercise may have to increase their time spent in the gym and subtract the time on the couch. While, this is possible, it requires far greater effort than declaring being athletic via ticking a box. Further, athletic ability maybe objectively assessed through a smartphone by monitoring speed of running/cycling and objectively testing behaviour. This is exactly what social psychology could benefit from.

Social Psychology's preferred method to study behaviour has morphed from observing behaviour in the real world to having participants complete self-report scales. This has not escaped the notice of researchers (Baumeister, Vohs, & Funder, 2007; Cialdini, 2009; Doliński, 2018). Their work charts the increasing absence of behaviour being studied in the landmark journal - *Journal of Personality and Social Psychology*. The proportions of papers that reported actual behaviour examination plummeted from a high of roughly 80% in 1976 to 2.6% in 2006 and 0.3% in 2017. It is worth noting that 2000-10 was designated the decade of behaviour by the American Psychological Association (Kendall, 2000)! Baumeister and colleagues in their paper postulated that the origins of this decline were due to multiple influences. First, the journal begun to require researchers to report multiple studies. Self-report methods would require a more manageable workload than behavioural methods. Second, ethics committees were suggested as having played a role. They would push psychologists to collect data that was less invasive and directed them towards methods that would ask people about their behaviour rather than monitor the behaviour of interest. Third, if a researcher used the same number of participants for an observational study as a questionnaire study they could have greater statistical confidence in the associated inferential statistics due to the nature of the data involved. This is because data returned from observing behaviour is frequently dichotomous or ordinal; however, self-report data is frequently interval (or at least treated as such regardless of the statistical appropriateness) and less participants are required as a result. If a researcher wants 80% confidence in the results of an experiment, they require different levels of evidence depending on the type of experiment. If they are studying a medium effect size in an experiment with two conditions, they would require either 64 participants if the data collected is interval or 107 if the data is dichotomous.

Interestingly, there is good evidence to suggest that calls for social psychologists to refocus their efforts on observing behaviour is not effective. Baumeister, Vohs, & Funder, (2007) attempted to do so (with a discrediting tone), but there was no change following the next time the journal was studied. In fact, rate of behaviour found in the *Journal of Personality and Social Psychology* may have become slightly worse (Doliński, 2018). However, this may have been due to the type of call made to researchers. Their attempt to shame researchers into good behaviour ignored some of the substantial incentives that exist, despite highlighting the incentives themselves. Nor, the reasonableness of using self-report data in science was acknowledged: for the researcher to be more efficient and ethical/non-invasive. However, these benefits are exactly what smartphones offer while also returning a much higher quality of data.

Calls for psychologists to begin to benefit from smartphones in their research have been made multiple times (Miller, 2012; Harari, et al., 2016). But unfortunately, psychologists' attention seems to centre on researching one aspect of the smartphone - the degree that smartphone use harms the mental and physical health of people. This was very predictable, smartphones are the most recent technology to be scrutinised by social psychologists for its potential negative impact on children and other marginalised groups in society (Orben, 2020) past focus of such concerns have included novels, radio, television (Bandura, Ross & Ross, 1963) and video games (Etchells, 2019). Some psychologists have conceptualised smartphone usage as negative because it reduces time spent other healthier pursuits, e.g. socialising, sleeping (Neuman, 1988). Others postulate that smartphones interfere with the healthy social interactions (Kushlev, & Dwyer, & Dunn, 2019). Indeed, this theory is not very original, as this concern has been documented 100 years prior to the development of the iPhone. In a 1906 magazine called *Punch* an illustration presented the question would wireless radio impact the family. The illustration captured a man and a woman sat together but listing into separate devices, the caption read: "These two figure are not communicating with one another. The lady is receiving an amatory message, and the gentleman some racing results" (Merchant, 2017, p. 44). The title of the image reads "Forecasts for 1907". The magazine puts forwards the concern that when people can get updates of interest remotely will they fail to attend to their relationships? But pretty damming evidence has shown that the engagement in a digital world is not necessary any more or less healthy than our analogue world. When reviewing the data from over 355,000 adolescent participants digital device usage and their wellbeing researchers concluded that "between digital technology use and adolescent wellbeing is negative and small, explaining at most 0.4% of the variation in well-being [and]… these effects are too small to warrant policy change" (Orben, & Przybylski, 2019, para. 1).

Beyond the above, there are multiple sensors included in the smartphone which hold promise for psychological research including location tracking sensors.

1.4.3 Data gathered by smartphones

Miller (2012) reviewed the sources of data that were available to psychological researchers. However, 10 years has passed since his article and his assortations can be reviewed with the benefit of seeing how technology has changed. Specifically, are there new unanticipated capabilities, were their heights anticipated for technological achievement reached and what predictions would not come to pass?

1.4.3.1 Data from Precursors to smartphone

Prior to smartphone research, mobile phones were used for psychological research. Researchers would send text messages to the participant with questionnaire items. These questions would be typically pointed at getting insights regarding recent time. For instance, Schnall and colleagues (2013) asked what health related questions did they have and what sources had they used to ask their questions. This method was simple, but allowed for questionnaires to be posed while participants were in an ecological setting and the recency of the subjects meant recall was far less of an issue than with typical questionnaires. But the re-programmability of these devices was limited and questionnaires needed to be administered over text messages. Surveys administered via text message were found to be responded to more reliably and more promptly than pen and paper alternatives (Berkman, Giuliani, & Pruitt, 2014). Past research has found that teenagers were willing to report accounts of their health via text messages (Schnall et al., 2013). It is argued (Mehl, Robbins, & große Deters, 2012) that with these ecological momentary assessments typically fails to provide data which is appropriately framed. The question is posed "how much less time does a person with a 4 on a social integration [out of 5] scale spend alone compared with a person with a 2?" (Mehl, Robbins, & große Deters, 2012, p. 414). Furthermore, is there consensus for what amount of time required to spend on social interactions in order to achieve a 5? Is there an interval difference between these features?

Electronically activated recorded (EAR) methodology is aimed at establishing the degree of socialising that an individual engages in (Mehl, Robbins, & große Deters, 2012). This methodology involves a participant having an audio recorder on their person when they go about their day. The device will record audio for a few seconds at regular intervals , later these recordings are used to make inferences about the participants' behaviour. From these recording a number of aspects will be inferred (Mehl, Gosling & Pennebaker, 2006): the type of activity the participant is engaging in, interaction with others, and emotional expression during this activity or interaction.

This methodology interestingly has demonstrated long ago that there is a trade-off between participant privacy and data quality. For the duration the period of recording that occurs during an EAR study linearly relates to the likelihood of mischaracterizing the degree of psychosocial activity that is occurring. A small snippet may simply catch a gap in conversation: type 2 error. Alternatively, a person may be ordering a coffee or having other mundane transactional conversation, this may not represent a meaningful socializing and cause a type 1 error. A similar relationship exists to the smartphone. Also, the accuracy of inferring activity and location seems low.

A next generation of EAR studies which offers a better balance of data quality and privacy may be a psych app. A psych app can run in the background and record the sounds which is occurring (possible in android), and then in real-time run an algorithm which recognizes speech and that of the participant specifically. Then every second (or other suitable duration) where speech is detected then is recorded. Then the sound is never recorded and deleted as a result. Therefore, there is no possibility for the content of the speech to be divulged to the researcher but the likelihood of the type 1 and 2 errors is minimised. There would however be concerns regarding the battery consumption of this app. Additionally, other sensors' location (GPS) and activity (accelerometer) may be better inferred.

1.4.3.2 Location data

GPS was used to record the location of the smartphone and this would return "±10 m for latitude and longitude, ±15 m for altitude, ±10 nanoseconds for time" (Miller, 2012, p. 225). GPS coverage is far greater and in rich nations there is up to a 5 meter accuracy. Yet, Miller's prediction that GPS precision could reach 1 meter was not necessarily inaccurate, such high level of accuracy is achievable. While this is typically not the case with smartphones, this is due not to signal from the satellites which can be used to measure within centimeters (Space Force, 2022), it is the limitations associated with the smartphone battery as simply the processing required for such accuracy would be too great. Additionally, Wi-Fi triangulation is available. The phone can source location from multiple separate Wi-Fi routers and then use this information to get accuracy of the person at about 1 meter (Android, 2022). This informs researcher not only which building they are in but researchers should be fully capable of knowing about the length of time that a participant spends in the each room of the house, given a layout design of the house.

Effective location tracking can provide a researcher with insights on where someone spends their time over a lengthy period, the duration of time that they spend in a locale, the adherence to a daily commute, the entropy of their movements, areas which are avoided and much more. Such metrics have been employed to make many different types of inferences about individuals' personality. For instance, the time dedicated to areas spent for socialising is larger in extraverts than introverts (Matz, & Harari, 2020). Interestingly, research into identifying personality from location tracking may suffer recruitment bias, likelihood to permit location sensing in a smartphone has been identified as being impacted by personality (Junglas, & Spitzmuller, 2006). Agreeable, conscientious and open to experience individuals are typically more willing to provide access to their smartphone location sensing. As a result, are only people who are open to experience, conscientious and agreeable applying to participate in studies that track location? Or is this no longer a methodological issue as attitudes may have been forced to change in a socially distant world? Many individuals have been pushed to use socially distant methods of consuming via hospitality sector mobile apps; some of which rely on location sensing (Mariano, 2020). Additionally, the wide use of apps that rely on location sensing have become much more common place for ordering food, hailing a taxi or even dating.

1.4.3.3 Visual input and recording data

Miller (2012) hypothesized that smartphones would accurately capture eye gaze. This has been attempted multiple times in last 10 years, but development and adoption of these devices has been fraught (Gvora, 2022, Stein, 2021). While, there are high quality solutions being developed of glasses which can achieve this with impressive results, glasses are being sold which track saccades without set up or calibration (Pupil labs, 2022). Additionally there are add-ons to VR and augmented reality (AR) set ups to allow for recording of saccades without setup being required. However, their core product is an expensive price tag currently associated with them of around £5000.

Visually capturing data regarding the surroundings has extensive potential. Building algorithms for computer vision is widely done. An object recognition algorithm that runs in the background could be employed. . Whenever a phone is in use or being held, recording of the environment and cataloguing what is in the presence (capable in android) can be done. Such abilities would be interesting in reporting the environment when phones are in use: Is the individual commonly surrounded by others, how frequent is the person around gym equipment, how frequently do they use public transport? Facial recognition could also be integrated into the analysis allowing for the identification of the social network. Interestingly, there may also be a potential for interaction between sensors and the visual inputs of the smartphone. A limitation of this suggestion is that images would only be captured when smartphones are to be used.

Other experiments are possible through using the smartphone camera and associated algorithms. A psych app could be designed that is used to capture the relative biometrics of an individual and then their food and drink intake then make estimates of what relative food would mean for their health.

1.4.3.4 External sensors

Miller (2012) commented multiple times on the possibilities of incorporating multiple sensors with the smartphone. While, Miller (2012) stated that biosensors would be common place either "wearable, implanted or injected" (p. 225) by 2025. This seems to be generally an over-estimation of the adoption of technology capable of measuring our biometrics.

One major roadblock is the health consequences associated with tech being implanted or injected. When invasive medical devices are implanted into an individual, they can potentially create a port of entry for pathogenic microorganisms, thereby increasing the risk of infection drastically (Safdar, Crnich, & Maki, 2001). Some examples of currently in use invasive medical devices include orthopedic or cardiac prostheses, vascular catheters, urinary catheters, and endotracheal tubes. Taking the example of prosthetic joint replacements, overall, 1%–1.5% of all patients who had hip or knee replacements suffered implant-related infections (Taylor & Webster, 2011). Moreover, this infection incidence rate is even higher when the device is implanted in immune-privileged sites, such as the central nervous system, the heart, and the eyes, where the immune response is less prominent. For instance, the frequency of infection is up to 27% with External Ventricular Drainage (a device inserted into the brain to directly monitor intracranial pressure) (Beer, Lackner, Pfausler, & Schmutzhard, 2008; Lozier, Sciacca, Romagnoli, & Connolly, 2002). Furthermore, such device-related infections are often tougher (more persistent) and total eradication is extremely hard to achieve due to the formation of biofilm.

Biofilm is a community of microorganisms that live in an extracellular matrix formed on the device surface, which protects them from antimicrobials and host immune attack(Jana Jass, Susanne Surman, & Walker, 2003; Taylor & Webster, 2011). Indeed, experiments demonstrated a five- to eight-fold increase in drug resistance to all antifungals relative to the same fungal pathogen in a non-biofilm state (Hawser & Douglas, 1995), the mortality is nearly doubled when infections are caused by biofilm-producing fungal pathogens versus non-biofilm producers (Vitalis et al., 2020). Besides this enormous risk of infection, other risks such as haemorrhage (bleeding), host rejection response (also called biofouling) to the implanted device, and pain or discomfort are all to be considered before committing to an invasive device (Harding & Reynolds, 2014). As such, inserting an external device into a person's body requires adequate justification, like monitoring health or replacing body functions, when less invasive routes are not available. It is also necessary to obtain regulatory body approval.

Hence, general scientific curiosity like Millers is generally not sufficient justification for exposing individuals to such risks. Potentially contributing to the skepticism is the scandals that have arisen with the development of medical devices such as blood monitoring devices designed by Therenos (Carreyou, 2019). Here, a startup company had engaged in wide scale fraud, including using medical devices to test for cancer when there was no justification to believe this was possible. Additionally, more recently, NeuralLink (2022) decided to develop a device that measures electrical activity on the brain from inside the skull. On advice from veterinarians 6 of the 23 monkeys in this experiment needed to be euthanized because of how the device had affected them (Ryan, 2022). Other reports have suggest that actually 15 out of 23 were killed due to extreme suffering (Graves, 2022).

However, electronic watches that capture biometrics are quite commonly used, at the end of 2018, 1 in 6 US adults had smart watches (Whitwam, 2019). This does limit the actual biometrics to heart rate sensor and blood oxygen sensor (Mills, 2020).

The smartphone does offer a very suitable platform for collecting and managing external devices. Such as the Electroencephalogram (EEG), portable versions of this device have been constructed which can run in tandem with the smartphone phone and psych app. Examples include EEGs (Pinho, Cerqueira, Correia, Sousa, & Dias, 2017; Boquete et al., 2012; Park, Myung, & Yoo, 2013), galvanic skin response (Navea, Buenvenida, & Cruz, 2019), but generally any sensor can be integrated with a smartphone. For example, smartphones are being used to sense the need for irrigation in crops using moisture detection (Jagüey, Villa-Medina, López-Guzmán, & Porta-Gándara, 2015).

However, what seems like it would have been surprising to Miller was the development of the internet of things (IoT). This allows for connection between external devices all over the internet. With a smart house, it would be capable for knowing the content of a participant's fridge (Samsung, 2022a), when a TV is switched on (Samsung 2022b), how regularly their house is heated, when it was occupied, when it clothes were washed, (Samsung, 2022c) and even how often they make toast (Deorsa, 2022) etc. But this of course is only an option when studying the very well off (especially with increasing electricity prices). Additionally, information about how frequent items are bought online including for weekly shopping (Barber, 2022) could be available.

The issue of the external sensors seems to raise issues of truly capturing what a smartphone is capable of, or what it is not. If data about a person or their actions is captured electronically and that individual has ownership of the data, this should be accessible to the smartphone and this may then be accessible to psychological research. Miller (2012) asserts that data from functional magnetic resonance imaging (FMRI) could not be captured by a psychological researcher because an FMRI will weight more than many cars (and many other reasons). But if the FMRI allows for the data to be shared with a psych app is this obstacle still a problem?

1.4.4 Examples of previous apps collecting data in ecologically valid settings

There has been multiple attempts to develop 'psych apps' These apps have been developed in order to research what can be inferred from smartphone data regarding experience of mental health illness and personality.

Much work has been done establishing what can be inferred about mental health from a user's smartphone. Experience of stress as established through reporting of stress scales across a two week period could be predicted with 72.28% accuracy by utilizing location data subsequently used to identify the weather in combination with Bluetooth and communication logs (Bogomolov, Lepri, Ferron, Pianesi, 2014). Sleep duration, variability in movement and amount of daily-movement could be utilised to inference experience of stress (Ben-Zeev, Scherer, Wang, Xie, & Campbell, 2015). Through establishing where a person spends time, and when they used their smartphone, Saeb and colleagues (2015) could predict with participants reporting of their sub-clinical depression (as established as a result of the patient health question-9; Kroenke, Spitzer & Williams, 2001) with 86.5% accuracy their experience of symptoms of depression within a healthy cohort of participants. Location can also be used to establish levels of social anxiety (Huang et al., 2016). By building machine learning models to understand experience of sub-clinical depression Ware and colleagues (2020) could very accurately predict self-reported experience of mental illness in non-clinical populations. Indeed, a meta-analysis recently found that smartphones may be useful for identifying experience of mental health problems (Weisel, Fuhrmann, Berking, Baumeister, Cuijpers,& Ebert, 2019). Indeed, another review found that smartphones can measure properties of behaviour relevant to mental health but the literature has yet to advance to demonstrate the ability for these apps to actually positively impact mental health (Aledavood, et al., 2019). The irregularity of a person's movements, the unpredictability of their movements and the distance travelled related to the individual's general well-being (Müller, Peters, Matz, Wang, & Harari, 2020). Indeed, the amount of activity and the degree that the person was moving related to their reported happiness (Lathia, Sandstrom, Mascolo, & Rentfrom, 2017).

Smartphones have also been employed to study personality. By monitoring a person for two separate weekends, the participants data (light data, noise level, battery level, accelerometer, call history, screen on/off, pedometer and location) could be used to predict personality with a high level of accuracy (between 66 and 71%; Khwaja, & Matic, 2019). While, data minimisation was attempted in this last study, extensive amount of data was captured from separate data sources. But this paper signifies the times when the participant's actions are a product of their own decisions - the weekend rather their action being influenced by other's choices (employer, school, etc.). Another study (Stachl, et al., 2017) showed that that reviewed the degree that an app usage related to multiple facets of personality. Participants who used their phone more to actively engage with apps (opposed to having the smartphone perform functions in the background like play music) were found to be less extraverted, conscientious and agreeable. Personality scores have been predicted from a combination of sensors such as Bluetooth and app usage (Chittaranjan, Blom, & Gatica-Perez, 2013).

1.5 Issues related to using smartphones in psychological research

Maintenance will always be an issue for software developers however it is a challenge which psychologists are also now having to deal with as part of their research. Beyond ensuring that code involved with analysis or stimulus creation remain functional to allow for replication, apps also pose similar concerns. For example, psych apps which are no longer available on the

Google Play store include: ContextSense (Chen, et al., 2014), Aware (Ferreira, Kostakos, & Dey, 2015) (source code is hosted publicly). Whereas Purple Robot (Saeb, et al., 2015), Funf in a Box (Aharony, et al., 2011), Soundsense (Lu, et al., 2009) & Studentlife (Wang et al., 2014) cannot be downloaded from app stores and code is not publicly accessible. Available apps that rely on servers are also particularly difficult to maintain. Even well-funded organizations like the New York Times do not find it practical to maintain an app that logged location and reports this to a server (New York Labs, 2011). Many experiments using a smartphone app for monitoring behavior do not make the app publicly available or the code and therefore it is very difficult to perform any sort of systematic review into unmaintained psych apps.

Laws (or rules) have been suggested which govern software maintenance (Lehman, 1980). Three of these laws were that software must continually update or it risks becomes irrelevant, software increases in complexity as it upgrades & the quality of the code depreciates as the development cycle continues. Indeed, when analysing the development of the Linux kernel (Israeli, & Feitelson, 2010) it was found that there was support for all of these laws. However, there were two laws which were not supported: that all updates to the code were small (occasional changes to code base were persuasive and profound) and changes to the organisation supporting the code change would be minor (the volunteer/open-source production of Linux meant that individuals would join and leave casually – this set up was far outside of what Lehman will have been exposed to). But importantly, the most supported law was that software updates were required to stay competitive. It may not be initially obvious why academic building psych-apps are subject to competition with other software, but this is because academics building apps for smartphones will always rely on the operating system, which itself is subject to regular upgrades.

Developers of psych apps therefore need to respond to how the operating system matures/changes there might be sufficient changes that means that the underlying software must be updated to keep it operational. In the Android system, an app must routinely be updated and published to the Google Play store to comply with the newest versions of the android system or the smartphones which use the newest versions of the android cannot install the app. Even if the app can be installed much functionality may be removed. The developers of the architecture are also required to judge which feature are worth maintaining based on popularity of the feature and this changes overtime (e.g. android architecture developers have done multiple U-turns for example when determining how androids can communicate to other devices via near field communication (Android, 2021c)). Therefore, apps developed need to make maintenance as simple as possible to minimise the cost to doing so, but future-proofing an app is very difficult.

Operating systems manage all interactions with a participant, the sensors, and third-party apps. The operating systems rules are subject to change and adapting in response to shifting cyber security concerns, legal regulation, world events (Covid pandemic; Apple, 2020), market pressures and many other factors. Designing the function of an operating system is a very difficult task, which requires maturation and slow growth. The immature operating system can be exploited for nefarious purposes at the expenses of the user. While platforms like Facebook & Android were immature, they typically provided great freedom to the developer at the expenses of the users (Cambridge Analytica is an example of this; Wylie; 2019). However, this resulted in widespread bad behaviour by the app developers, they would typically not consider how their apps was harming how the platform was operating (Android Developer, 2017) and would prompt restrictions being introduced to the operating system. Apple's iPhone operating system may have extracted lessons from developing computers for over 20 years at the point of the iPhone development. Perhaps for this reason, we do not see the same maturation process that Google and Facebook needed to undergo with regards to developer behaviour.

Operating systems therefore have to several roles to play. One important role is to ensure the health of the phone overall. How the operating system achieves this adapts over time. The Android system began to restrict the degree that a program can run in the background (Android Developer, 2017). This was to overcome the issue of apps running excessively in the background and significantly reducing the battery life of the apps and therefore keep the android smartphones competitive against iPhones and Blackberrys. The operating system does metaphorically take on the role of an immune system in a human body. An android app prior to 2017 could have acted like a parasitic tape worm and consume the resources without providing anything of value to the host. To combat this, Android implemented something called doze mode that pauses all processes running in the background unless it is essential for the function of the app. If it was essential, then the app would need to inform the user that operations were being carried out & allow the user to cancel the operations. Later, more sophisticated battery resource allocation was introduced subsequently, apps which the user engaged with regularly would be able to run for longer in the background. But the doze mode remained. But this is not the only issue for interacting with smartphone operating systems. Continuing with the immunity metaphor, the smartphone operating system was concerned that a virtual virus and therefore stopped passive across Bluetooth connectivity and thereby increased the likelihood of users acquiring a real virus (covid-19; Sabbagh, & Hern, 2020).

Apple had particular problematic levels of functioning because the iOS stopped Bluetooth being used by apps running in the background (Sabbagh & Hern, 2020) for covid-19 track and trace apps. A hack to this problem was implemented, where the app would wake itself up if encountered with another user. The hope was that the Apple smartphones would be kept awake sufficiently by Android phones being nearby and waking up the Apple phone. This did not happen, only 4% of the time did the apple app function properly. The same issue undermined android as 75% of the time did it work. Under normal circumstances having Bluetooth run in the background is a needless security risk as multiple phones based cyber-attacks can occur via Bluetooth (Melamed, 2018). However, android and apple released a system that was exempt to the usual operating system provided a registered healthcare provider published the app this therefore could radically increase the accuracy for Covid track and trace apps (Sabbagh & Hern, 2020). The limitations of the operating system plagues entire countries never mind small research labs. Regardless, we will explore if there is some method that apps can be better designed to best avoid clashing with the smartphone operating systems.

Operating systems are clearly a significant factor in the development of apps. To ensure this thesis remains manageable, we will be ensuring all construction of software relies upon one operating system – Android. Androids offer far more affordances to developers than iPhones. Androids allow for almost free distribution across their play store (Google, 2021c) and do not suffer from a notoriously strict and secretive app review procedure (Leswing, 2021). Android software can also easily be distributed outside of the play store but there are extensive limitations for this on an iPhone like limiting the number of apps which can be developed (Apple Developer, 2017). Android represent 80% of the smartphone market (Keil et al., 2020) therefore more participants can be reached via Androids.

Indeed, multiple other issues exist within smartphone psychological research including: developing applications which are overly invasive (commonly done in this research area), software that is participant-unfriendly and limited sharing of technical solutions, etc. However, these challenges within psychological research are specifically related to the culture adopted by researchers. Some of these challenges are beyond the remit of this thesis. Instead, I intend to focus on overcoming the technical limitations for smartphone based psychological research. However, providing high quality methodological guidelines and publicly available open-source software will also peripherally contribute to improvements in psychological science more generally.

1.6 Aims of the Thesis

This thesis aims to :

1. Develop psych apps capable of being used by psychologists. These must meet the following criteria: One - There must be a very low or no barrier to entry to employ the software. The adoption of websites for psychological research did not truly begin until Amazon Mechanical Turk reduced the technical skills required to set up the system (see above). Two - many psych apps have historically been too expensive to maintain as they relied upon servers. Build software which does not need to rely upon such infrastructure to function. Three - to ensure that the apps can be customized for a range of experimentation and that this process is as straightforward as possible. Develop this software for two separate types of digital traces:
    a. Develop software capable of effectively capturing the context of behaviour.
    b. Develop software that captures behaviour.
2. Establish if there is a theoretical grounding for the use of psych apps relative to conventional methodologies
3. Employ psych apps in empirical research

1.7 Thesis outline

Collectively, this thesis will enhance our understanding of digital traces by first developing apps that are suitable for psychological research. Through this process, I intend to outline an effective method for developing apps for psychological research and provide working code that can be developed further. The thesis will then document the findings of subsequent psychological research that has utilised these apps.

Part 1 – Building Apps

Chapter 2 – Documenting context [published in *Behavior Research Methods*]

This chapter outlines the development of an app that was developed and published for Android smartphones to log location by using multiple sources. Past research utilising location sensing is also reviewed. This represents a first attempt in developing an app which can be readily used by psychologists for passively logging rich behavioural data.

Chapter 3 – Documenting digital behaviour [published in *Behavior Research Methods*]

Technology usage remains a controversial topic in cyber psychology. I take the lessons learnt from developing the previous two apps and use this to build an app which is highly customisable and can deliver high quality insights into how participants are using their smartphones. Ultimately, providing a new tool which may contribute to ongoing debates regarding the impacts of general and specific smartphone use.

Part 2 – Using Apps

Chapter 4 - Subjective reports overstate the relationship between screen time and mental health [published in the *International Journal of Human-Computer Studies*]

Previous methodologies (self-report) have implied that there is a link between general smartphone use and negative health outcomes. We use objective measures of screen usage via the apps previously developed and demonstrate that the self-report measures lead researchers to falsely conclude that there is a strong link between screen time and poor health.

Chapter 5 – Quantifying smartphone 'use': Choice of measurement impacts relationships between 'usage' and health [published in *Technology, Mind and Behavior*]

This paper reports on evidence which suggests when using smartphone scales to identify using a smartphone on physical and mental health are overestimated when unobjective measurements of smartphone use are employed. Additionally, the methodology expands beyond the use of Apple screen time to give a more in-depth account of smartphone usage by using a version of the android app developed previously.

Chapter 6 - Conclusion

The conclusion will review the novel contributions of the thesis and combine these in a demonstration of how smartphones can supply data about context/environment, biometrics and behaviour of individuals simultaneously. After reviewing limitations of the thesis, I will make the recommendations for research in the field and suggest how further research should be conducted into “mini-psych apps” to overcome existing limitations.

References

Aaronson, D., Grupsmith, E., & Aaronson, M. (1976). The impact of computers on cognitive psychology. Behavior Research Methods & Instrumentation, 8(2), 129-138.

Aharony, N., Pan, W., Ip, C., Khayal, I., & Pentland, A. (2011, September). The social fMRI: measuring, understanding, and designing social mechanisms in the real world. In Proceedings of the 13th international conference on Ubiquitous computing (pp. 445-454).

Aledavood, T., Torous, J., Hoyos, A. M. T., Naslund, J. A., Onnela, J. P., & Keshavan, M. (2019). Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Current psychiatry reports, 21(7), 49.

Amazon (2022). Mechanical Turk. Retrieved from https://www.mturk.com/

Andreani, W., & Ying, Y. (2019). " PowPow" interactive game in supporting English vocabulary learning for elementary students. *Procedia Computer Science*, *157*, 473-478.

Android (2021a) Intents [Website]. Retrieved from 'https://developer.android.com/reference/android/content/Intent'

Android (2021b). Permissions on Android [Website]. Retrieved from 'https://developer.android.com/guide/topics/permissions/overview'

Android (2021c) Near field communication overview [Website]. Retrieved from: 'https://developer.android.com/guide/topics/connectivity/nfc'

Android (2022) Wi-Fi location: ranging with RTT [Website]. Retrieved from: 'https://developer.android.com/guide/topics/connectivity/wifi-rtt'

Apple (2020). Apple Releases new COVID-19 app and website based on CDC guidance [Website]. Retrieved from https://www.apple.com/newsroom/2020/03/apple-releases-new-covid-19-app-and-website-based-on-CDC-guidance/

Apple Developer (2017). The maximum number of apps for free development profiles has been reached [Website]. Retrieved from "https://developer.apple.com/forums/thread/49304"

Arnold, H. J. (1976). Effects of performance feedback and extrinsic reward upon high intrinsic motivation. Organizational Behavior & Human Decision Processes, 17, 275-288.

ArtiFact (2022) ArtiFact. Retrieved from: https://artifacts.ai/

Averill, C. (2022). A Brief Analysis of the Apollo Guidance Computer. *arXiv preprint arXiv:2201.08230*.

Azzouz Boudadi, N., & Gutiérrez-Colón, M. (2020). Effect of Gamification on students' motivation and learning achievement in Second Language Acquisition within higher education: a literature review 2011-2019. *The EuroCALL Review*, *28*(1), 40-56..

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012, June). Personality and patterns of Facebook usage. In Proceedings of the 4th annual ACM web science conference (pp. 24-32).

Bailey, A. W., & Hungenberg, E. (2019). Psychogeography of a Marathon Runner: An Exploratory EEG Approach. Journal of Sport Behavior, 42(4), 415-440.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences, 106(51), 21484-21489.

Bandura, A., Ross, D., & Ross, S. A. (1963). Imitation of film-mediated aggressive models. The Journal of Abnormal and Social Psychology, 66(1), 3.

Barberm R., (2022). Online Shopping Statistics & Trends in 2022. Retrieved from 'https://www.cloudwards.net/online-shopping-statistics/'

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? Perspectives on psychological science, 2(4), 396-403.

BBC (2021). Epic's bid to sue Apple over Fortnight in UK rejected [Online Newspaper article]. Retrieved from: 'https://www.bbc.co.uk/news/technology-56156799'

Beer, R., Lackner, P., Pfausler, B., & Schmutzhard, E. (2008). Nosocomial ventriculitis and meningitis in neurocritical care patients. J Neurol, 255(11), 1617-1624. doi:10.1007/s00415-008-0059-8

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatric rehabilitation journal, 38(3), 218.

Bergstrom, C. T., & West, J. D. (2020). Calling bullshit: the art of skepticism in a data-driven world. Random House.

BHaptics (2021) TactSuit X16. [Website] Retrieved from: "https://www.bhaptics.com/shop/tactsuit-x16"

Bhattacharjee, Y. (2013). The mind of a con man. The New York Times, 28.

Birnbaum, M. H. (2001). *Introduction to behavioral research on the Internet*. Pearson College Division.

Birnbaum, M. H. (2004). Human research and data collection. *Annu. Rev. Psychol*, *55*, 803-32.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. Psychological bulletin, 117(2), 187.

Blum B. (2018) The Lifespan of a Lie. The most famous psychology study of all time was a sham. Why can't we escape the Stanford Prison Experiment? [Website]. Retrieved from https://gen.medium.com/the-lifespan-of-a-lie-d869212b1f62.

Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A. (2014, November). Daily stress recognition from mobile phone data, weather conditions and individual traits. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 477-486).

Bohannon, J. (2016). Mechanical Turk upends social sciences. Retrieved from https://www.science.org/doi/full/10.1126/science.352.6291.1263

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295-298.

Boquete, L., Ascariz, J. M. R., Cantos, J., Barea, R., Miguel, J. M., Ortega, S., & Peixoto, N. (2012). A portable wireless biometric multi-channel system. Measurement, 45(6), 1587-1598.

Bowler, D. M. (1987). Using an Apple II to control a two-field projection tachistoscope. Behavior Research Methods, Instruments, & Computers, 19(4), 377-379.

Brownlee, J. (2021). How to Code a Neural Network with Backpropagation in Python (from scratch). Retrieved from https://machinelearningmastery.com/implement-backpropagation-algorithm-scratch-python/

Buss, D. H. (1972). The effect of tachistoscopic training on reading comprehension.
Cahalan, S. (2019). The great pretender: The undercover mission that changed our understanding of madness. Hachette UK.

Cahalan, S. (2019). *The great pretender: The undercover mission that changed our understanding of madness*. Hachette UK.

Capelle-Blancard, G., & Coulibaly, D. (2011). Index trading and agricultural commodity prices: A panel Granger causality analysis. *International Economics*, *126*, 51-71.

Carpenter, S. (2012). Harvard psychology researcher committed fraud, US investigation concludes. Science, 6.

Carreyrou, J. (2019). Bad blood. Larousse.

Cattell, R. B. (1958). What is" objective" in" objective personality tests?". Journal of Counseling Psychology, 5(4), 285.

Cattell, R. B. 1952. Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist. New York: Harper.

Chen, Z., Chen, Y., Hu, L., Wang, S., Jiang, X., Ma, X., ... & Campbell, A. T. (2014, September). ContextSense: unobtrusive discovery of incremental social context using dynamic bluetooth data. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (pp. 23-26).

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. Personal and Ubiquitous Computing, 17(3), 433-450.

Chorley, M. J., Whitaker, R. M., & Allen, S. M. (2015). Personality and location-based social networks. Computers in Human Behavior, 46, 45-56.

Cialdini, R. B. (2009). We have to break up. Perspectives on psychological science, 4(1), 5-6.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., ... & Nowak, A. (2012). Manifesto of computational social science. The European Physical Journal Special Topics, 214(1), 325-346.

de Gelder, B., Kätsyri, J., & de Borst, A. W. (2018). Virtual reality and the new psychophysics. British Journal of Psychology, 109(3), 421-426.

Deoras, S., (2022). First ever IoT device – "The Internet Toaster" [Website]. *Analytics in* diamag. Retrieved from 'https://analyticsindiamag.com/first-ever-iot-device-the-internet-toaster/'

Detour (2021) Detour [Software]. Retrieved from: 'https://play.google.com/store/apps/details?id=com.detour.detour'

Dhar, P. (2020) The most emotional deaths in gaming. *Looper*. Retrieved from https://www.looper.com/193560/the-most-emotional-deaths-in-gaming/

Doliński, D. (2018). Is psychology still a science of behaviour?. Social Psychological Bulletin, 13(2), 1-14.

Dumont, F. (2010). A history of personality psychology: Theory, science, and research from Hellenism to the twenty-first century. Cambridge University Press.

Echonest (2021). Echonest [Software]. Retrieved from 'blog.echonest.com'

Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. SIAM review, 21(4), 460-480.

Epic Games (2021). Fortnight [Software] Retrieved from: 'https://www.epicgames.com/fortnite/en-US/home'

Etchells, P. (2019). Lost in a good game: Why we play video games and what they can do for us. Icon Books.

Facebook (2021a) Facebook [Software]. Retrieved from: 'https://en-gb.facebook.com/'

Facebook (2021b). Messenger [Software]. Retrieved from: 'https://www.messenger.com/'

Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media. PloS one, 10(11), e0142390.

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: mobile context instrumentation framework. Frontiers in ICT, 2, 6.

FourSquare (2021) Foursquare [Website]. Retrieved from: 'https://foursquare.com/'

Funk, S. (2006). Netflix Update: Try this at Home [Website]. Retrieved from: 'https://sifter.org/~simon/journal/20061211.html'

Funk, S., Salathé, M., & Jansen, V. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of the Royal Society Interface, 7(50), 1247-1256.
Funk, S., Salathé, M., & Jansen, V. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of the Royal Society Interface, 7(50), 1247-1256.
Gaggioli, A. (2001). Using virtual reality in experimental psychology. Towards Cyberpsychology, 157-174.

Google (2021b) Google Timeline [website]. Retrieved from; 'https://google.com/maps/timeline'

Google (2021a). Google Maps – Navigate & Explore [Software]. Retrieved from: 'https://play.google.com/store/apps/details?id=com.google.android.apps.maps'

Google. (2021c). Play Store [Website]. Retrieved from: "https://play.google.com/store"
Google. (2021c). Play Store [Website]. Retrieved from: "https://play.google.com/store"
Graves, W., (2022). "Extreme Suffering": 15 of 23 Monkeys with Elon Musk's Neuralink Brain Chips reportedly died [Website]. *Yahoo Entertainment*. Retrieved from https://www.yahoo.com/entertainment/extreme-suffering-15-23-monkeys-205653491.html

Grindr (2021) Grindr [Software]. Retrieved from: 'https://play.google.com/store/apps/details?id=com.grindrapp.android'

Grusky, D. B., Western, B., & Wimer, C. (Eds.). (2011). The great recession. Russell Sage Foundation.

Guerrilla Games (2017). [Software]. Horizon Zero dawn.

Gvora, J., (2022). Google Glass: What happened to the Futuristic smart glasses? *Screen Rant*. Retrieved from https://screenrant.com/google-glass-smart-glasses-what-happened-explained/
H&M (2021). H&M [Software]. Retrieved from: 'https://play.google.com/store/apps/details?id=com.hm.goe&hl=en_GB&gl=US'

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

Harding, J. L., & Reynolds, M. M. (2014). Combating medical device fouling. Trends Biotechnol, 32(3), 140-146. doi:10.1016/j.tibtech.2013.12.004

Harris, K., & Vazire, S. (2016). On friendship development and the Big Five personality traits. Social and Personality Psychology Compass, 10(11), 647-667.

Hawser, S. P., & Douglas, L. J. (1995). Resistance of Candida albicans biofilms to antifungal agents in vitro. Antimicrob Agents Chemother, 39(9), 2128-2131. doi:10.1128/AAC.39.9.2128

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29-29.

Hodges, A. (2012). Alan Turing: the enigma. Random House.

Hollander, M. M., & Turowetz, J. (2017). Normalizing trust: Participants' immediately post-hoc explanations of behaviour in M ilgram's 'obedience' experiments. *British Journal of Social Psychology*, *56*(4), 655-674.

Hollemans, M. (2018). How fast is my model?. Retrieved from https://www.machinethink.net/blog/how-fast-is-my-model/

Huang, Y., Xiong, H., Leach, K., Zhang, Y., Chow, P., Fua, K., ... & Barnes, L. E. (2016, September). Assessing social anxiety using GPS trajectories and point-of-interest data. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 898-903).

Isen, A. M., Shalker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? Journal of Personality & Social Psychology, 36, 1-12.

Israeli, A., & Feitelson, D. G. (2010). The Linux kernel as a case study in software evolution. Journal of Systems and Software, 83(3), 485-501.

Jagüey, J. G., Villa-Medina, J. F., López-Guzmán, A., & Porta-Gándara, M. Á. (2015). Smartphone irrigation sensor. IEEE Sensors journal, 15(9), 5122-5127.

Jana Jass, Susanne Surman, & Walker, J. (2003). Medical Biofilms: Detection, Prevention and Control: John Wiley & Sons, Ltd.

Jensen-Campbell, L. A., Adams, R., Perry, D. G., Workman, K. A., Furdella, J. Q., & Egan, S. K. (2002). Agreeableness, extraversion, and peer relations in early adolescence: Winning friends and deflecting aggression. Journal of Research in Personality, 36(3), 224-251.

Jobs, S., (2007). Steve Jobs Introducing The iPhone at MacWorld. Retrieved from 'https://www.youtube.com/watch?v=x7qPAY9JqE4'

Jones, M. B. (1984). Video games as psychological tests. Simulation & Games, 15(2), 131-157.

Jones, M. B., Kennedy, R. S., & Bittner, A. C. (1981). A video game for performance testing. American Journal of Psychology, 94,143-152.

Junglas, I., & Spitzmuller, C. (2006, June). Personality traits and privacy perceptions: an empirical study in the context of location-based services. In 2006 International Conference on Mobile Business (pp. 36-36). IEEE.

K. Mariano (2020). Why digital apps must be part of hospitality industry's 'new normal' [Online Magazine]. Travel daily media. Retrieved from: https://www.traveldailymedia.com/digital-apps-hospitality-industry-new-normal

Karsai, M., Kivelä, M., Pan, R. K., Kaski, K., Kertész, J., Barabási, A. L., & Saramäki, J. (2011). Small but slow world: How network topology and burstiness slow down spreading. Physical Review E, 83(2), 025102.

Keil, T. F., Koschate, M., & Levine, M. (2020). Contact Logger: Measuring everyday intergroup contact experiences in near-time. Behavior Research Methods, 1–19.
Kendall, P. C. (2000). The decade of behavior. Journal of Consulting and Clinical Psychology, 68(1), 3.

Khwaja, M., & Matic, A. (2019, September). Personality is revealed during weekends: Towards data minimisation for smartphone based personality classification. In IFIP Conference on Human-Computer Interaction (pp. 551-560). Springer, Cham.

Kitase, Y. (1997). [Software] Final Fantasy VII. *Square Enix*

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the national academy of sciences, 110(15), 5802-5805.

Kumar, V. (2020). Top 10 Fastest Supercomputers in the world to watch in 2020. *Analytics insight*. Retrieved from https://www.analyticsinsight.net/top-10-fastest-supercomputers-world-watch-2020/#:~:text=Summit%2C%20a%20supercomputer%20developed%20by%20IBM%20for%20use,current%20LINPACK%20benchmark%20is%20clocked%20at%20148.6%20petaFLOPS.

Kushlev, K., Dwyer, R., & Dunn, E. W. (2019). The social price of constant connectivity: Smartphones impose subtle costs on well-being. Current Directions in Psychological Science, 28(4), 347-352. nonprogrammers. Behavior Research Methods, Instruments, & Computers, 16(1), 25-27.

Kusuma, G. P., Wigati, E. K., Utomo, Y., & Suryapranata, L. K. P. (2018). Analysis of gamification models in education using MDA framework. Procedia Computer Science, 135, 385-392.

La Scaleia, B., Ceccarelli, F., Lacquaniti, F., & Zago, M. (2020). Visuomotor interactions and perceptual judgments in virtual reality simulating different levels of gravity. Frontiers in bioengineering and biotechnology, 8.

Lapin, T. (2018). Framed Stanford Prison Experiment was a fraud, scientist says. *The New York Post*. Retrieved from https://nypost.com/2018/06/14/famed-stanford-prison-experiment-was-a-fraud-scientist-says/

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. PloS one, 12(1), e0160589.

Lehman, M. M. (1980). Programs, life cycles, and laws of software evolution. Proceedings of the IEEE, 68(9), 1060-1076.

Lerman, K., Ghosh, R., & Surachawala, T. (2012). Social contagion: An empirical study of information spread on Digg and Twitter follower graphs. arXiv preprint arXiv:1202.3162.

Leroy, A., & Cheron, G. (2020). EEG dynamics and neural generators of psychological flow during one tightrope performance. Scientific reports, 10(1), 1-13.

Leswing, K., (2021). Inside Apple's team that greenlights iPhone apps for the App Store. [Online news article]. CNBC. Retrieved from: https://www.cnbc.com/2019/06/21/how-apples-app-review-process-for-the-app-store-works.html.

Lewis, M. (2004). Moneyball: The art of winning an unfair game. WW Norton & Company.

Lewis, M. (2004). Moneyball: The art of winning an unfair game. WW Norton & Company.

Lewis, M. (2016). The undoing project: A friendship that changed the world. Penguin UK.

Lewis, M. (2016). The undoing project: A friendship that changed the world. Penguin UK.

Linkenauger, S. A., Bülthoff, H. H., & Mohler, B. J. (2015). Virtual arm 's reach influences perceived distances but only after experience reaching. Neuropsychologia, 70, 393-401.

Lisetti, C. L., & Rumelhart, D. E. (1998, May). Facial Expression Recognition Using a Neural Network. In *FLAIRS Conference* (pp. 328-332).

London, L., (2020, March 23). Forbes. Integrated whale media investments. Retrieved from: https://www.forbes.com/sites/lelalondon/2020/03/23/in-self-isolation-filter-dysmorphia-and-beauty-filters-will-threaten-our-mental-health/?sh=222ed1163831

Lozier, A. P., Sciacca, R. R., Romagnoli, M. F., & Connolly, E. S., Jr. (2002). Ventriculostomy-related infections: a critical review of the literature. Neurosurgery, 51(1), 170-181; discussion 181-172. doi:10.1097/00006123-200207000-00024

Lu, H., Pan, W., Lane, N. D., Choudhury, T., & Campbell, A. T. (2009, June). Soundsense: scalable sound sensing for people-centric applications on mobile phones. In Proceedings of the 7th international conference on Mobile systems, applications, and services (pp. 165-178).

Lykins, D., (2019). The 10 Most Heartbreaking Moments in Final Fantasy, Ranked. *The Gamer*. Retrieved from https://www.thegamer.com/final-fantasy-franchise-saddest-moments-ranked/

Mackey, T. K., Shah, N., Miyachi, K., Short, J., & Clauson, K. (2019). A framework proposal for blockchain-based scientific publishing using shared governance. Frontiers in Blockchain, 2, 19.

MapMyFitness (2021) Map my Run [Software]. Retrieved from: 'https://play.google.com/store/apps/details?id=com.mapmyrun.android2'

Marshall, J. C. (1977). Minds, machines and metaphors. Social studies of science, 7(4), 475-488.

Matz, S. C., & Harari, G. M. (2020). Personality–place transactions: Mapping the relationships between Big Five personality traits, states, and daily places. Journal of Personality and Social Psychology.

McCarthy, P. (2017). Infograhic: The growth of computer processing power?. Retrieved from https://www.offgridweb.com/preparation/infographic-the-growth-of-computer-processing-power/

Melamed, T. (2018). An active man-in-the-middle attack on bluetooth smart devices. Safety and Security Studies, 15, 2018.

Merchant, B. (2017). The one device: The secret history of the iPhone. Hachette UK.

Mi, Q., Keung, J., Mei, X., Xiao, Y., & Chan, W. K. (2018, July). A gamification technique for motivating students to learn code readability in software engineering. In 2018 International Symposium on Educational Technology (ISET) (pp. 250-254). IEEE.

Mickel, C. (1982). Innovative projects earning psychologists spots on hospital health care teams. American Psychologist, 37, 1350-1354

MightySignal (2021). AppMonsta [Software]. Retrieved from: 'https://appmonsta.com/?gclid=EAIaIQobChMIsLKnroCm7wIVBqWyCh1LnwAmEAAYASAAEgIYL_D_BwE'

Milgram, S., & Gudehus, C. (1978). Obedience to authority.

Miller, B. (2011) Money ball [Film]. Columbia Pictures.

Miller, G. (2012). The smartphone psychology manifesto. Perspectives on psychological science, 7(3), 221-237.

Mills, M., (2020) Sensors of an Apple Watch: All that it has explained [Website]. Retrieved from: https://itigic.com/sensors-of-an-apple-watch-all-that-it-has-explained/

Morris, J. J., & Alam, P. (2012). Value relevance and the dot-com bubble of the 1990s. The Quarterly Review of Economics and Finance, 52(2), 243-255.

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the Relationships between Mobility Behaviours and Indicators of Subjective Well–Being Using Smartphone–Based Experience Sampling and GPS Tracking. European Journal of Personality, 34(5), 714-732.

Musch, J., & Reips, U. D. (2000). A brief history of Web experimenting. In Psychological experiments on the Internet (pp. 61-87). Academic Press.

Nagar, P., & Sethia, D. (2019, January). Brain mapping based stress identification using portable eeg based device. In 2019 11th International Conference on Communication Systems & Networks (COMSNETS) (pp. 601-606). IEEE.

Nakamoto, S. (2008). Bitcoin whitepaper. *URL: https://bitcoin. org/bitcoin. pdf-(: 17.07. 2019)*.

Navea, R. F., Buenvenida, P. J., & Cruz, C. D. (2019, November). Stress detection using galvanic skin response: An android application. In journal of physics: Conference series (Vol. 1372, No. 1, p. 012001). IOP Publishing.

Netflix (2009). Netflix Prize [Website]. Retrieved from: 'https://www.netflixprize.com/'

Netflix (2021) Netflix [Website]. Retrieved from: 'https://netflix.com'

Neuman, S. B. (1988). The displacement effect: Assessing the relation between television viewing and reading performance. Reading research quarterly, 414-440.

NeuralLink (2022). NeuralLink [Website] Retrieved from 'https://neuralink.com/'
New York Labs (2011). OpenPaths [website/software]. Retrieved from: https://nytlabs.com/projects/openpaths.html

Noë, N., Whitaker, R. M., Chorley, M. J., & Pollet, T. V. (2016). Birds of a feather locate together? Foursquare checkins and personality homophily. Computers in Human Behavior, 58, 343-353.

Nomura, T., Hamaguchi, N., & Toriyama, M. (2020). [Software] Final Fantasy VII Remake. *Square Enix*

Norris, D. (1984). A computer-based programmable tachistoscope for nonprogrammers. Behavior Research Methods, Instruments, & Computers, 16(1), 25-27.

O'Dwyer, D. & Jayne, J. (2017) The making of Horizon Zero Dawn. Film. Noclip
Orben, A. (2020). The Sisyphean cycle of technology panics. Perspectives on Psychological Science, 15(5), 1143-1157.

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. Nature Human Behaviour, 3(2), 173-182.

Park, H., Myung, B., & Yoo, S. K. (2013, February). Power consumption of wireless EEG device for BCI application: Portable EEG system for BCI. In 2013 International Winter Workshop on Brain-Computer Interface (BCI) (pp. 100-102). IEEE.

Paul, K., & Sweney, M. (2020) Apple and Google remove Fortnite video game from app stores. [Online newspaper article]. Retrieved from: https://www.theguardian.com/technology/2020/aug/13/fortnite-maker-lawsuit-apple-app-store-removal

Pinho, F., Cerqueira, J., Correia, J., Sousa, N., & Dias, N. (2017). myBrain: a novel EEG embedded system for epilepsy monitoring. Journal of medical engineering & technology, 41(7), 564-585.

Pokemon (2021) Pokemon Go [Software] Retrieved from: 'https://play.google.com/store/apps/details?id=com.nianticlabs.pokemongo'

Pupil Labs (2022) Pupil Labs [Website]. Retrieved from 'https://pupil-labs.com/'

Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*, *31*(1-2), 72-89..

Rawendy, D., Ying, Y., Arifin, Y., & Rosalin, K. (2017). Design and development game Chinese language learning with gamification and using mnemonic method. *Procedia Computer Science*, *116*, 61-67.

Ritter, S., Barrett, D. G., Santoro, A., & Botvinick, M. M. (2017, July). Cognitive psychology for deep neural networks: A shape bias case study. In International conference on machine learning (pp. 2940-2949). PMLR.

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, *179*(4070), 250-258.

Roulin, N., & Levashina, J. (2016). Impression management and social media profiles. In Social media in employee selection and recruitment (pp. 223-248). Springer, Cham.

Russo, D., W., (2016). 10 Most Heartbreaking Video Game deaths of All Time. *Screen Rant*. Retrieved from https://screenrant.com/best-most-heartbreaking-video-game-deaths-ever-all-time/#:~:text=%2010%20Most%20Heartbreaking%20Video%20Game%20Deaths%20Of,up%20copies%20of%20all-out%20splatterfests%20%28like...%20More%20

Ryan, H., (2022) Elon Musk's Nueralink confirms monkeys died in project, denies animal cruelty claims [Website]. *CNN.* Retrieved from https://edition.cnn.com/2022/02/17/business/elon-musk-neuralink-animal-cruelty-intl-scli/index.html

Sabbagh, D., & Hern. A (2020). UK abandons contact-tracing app for Apple and Google model. The Guardian. Retrieved from: https://www.theguardian.com/world/2020/jun/18/uk-poised-to-abandon-coronavirus-app-in-favour-of-apple-and-google-models

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. Journal of medical Internet research, 17(7), e175.

Safdar, N., Crnich, C. J., & Maki, D. G. (2001). Nosocomial Infections in the Intensive Care Unit Associated with Invasive Medical Devices. Curr Infect Dis Rep, 3(6), 487-495. doi:10.1007/s11908-001-0085-5

Samsung (2022a). RB31 Fridge Freezer with Digital inverter Technology [website]. Retrieved from 'https://www.samsung.com/uk/refrigerators/bottom-mount-freezer/308l-metal-graphite-rb31fdrndsa-eu/?_gl=1*1kndz1a*_up*MQ..&gclid=f41fbc6bc5de16725734df608fc7b3db&gclsrc=3p.ds'

Samsung (2022b). Smart TV [website]. Retrieved from https://www.samsung.com/us/tvs/smart-tv/connect-your-phone/#:~:text=Just%20tap%20your%20smartphone%20on%20the%20TV%20and,to%20view%20it%20on%20the%20big%20screen.%202

Samsung (2022c). Smart Home with IoT devices [Website]. Retrieved from 'https://www.samsung.com/levant/tvs/smart-tv/smart-home-with-iot-devices/'

Sharma, M., Kacker, S., & Sharma, M. (2016). A brief introduction and review on galvanic skin response. Int J Med Res Prof, 2, 13-17.

Silver, N. (2012). The signal and the noise: the art and science of prediction. Penguin UK.

Silver, N. (2012). The signal and the noise: the art and science of prediction. Penguin UK.

Singh, A., Click, K., Parizi, R. M., Zhang, Q., Dehghantanha, A., & Choo, K. K. R. (2020). Sidechain technologies in blockchain networks: An examination and state-of-the-art review. Journal of Network and Computer Applications, 149, 102471.

Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. Annual review of psychology, 57, 529.

Snow Inc. (2021a). B612 -Free selfe Camera, Photo Editor & Video App. [10.2.11]. Play Store. Retrieved from: https://play.google.com/store/apps/details?id=com.linecorp.b612.android&hl=en_GB&gl=US

Snow Inc. (2021b). B612 -Free selfe Camera, Photo Editor & Video App. [10.2.13]. App Store. Retrieved from: https://apps.apple.com/co/app/b612-camera-photo-video-editor/id904209370

Sony (1994). [Hardware] PlayStation.

Sony (2000). [Hardware] PlayStation 2.

Sony (2006). [Hardware] PlayStation 3.

Sony (2013). [Hardware] PlayStation 4.

Sony (2020). [Hardware] PlayStation 5.

Space Force (2022). *GPS Accuracy*. Retrieved from ‘https://www.gps.gov/systems/gps/performance/accuracy/#:~:text=Recent%20FAA%20data%20shows%20their%20high%20quality%2C%20single-frequency,%2810.5%20MB%20PDF%29%20Return%20to%20top%20of%20page’

Spotify (2021). Spotify [Software]. Retrieved from: ‘www.spotify.com’

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., ... & Bühner, M. (2017). Personality traits predict smartphone usage. European Journal of Personality, 31(6), 701-722.

Stein, S., (2021) Facebook’s Ray-Ban smart glasses are coming: what to expect [Website]. *CNET*. Retrieved from https://www.cnet.com/tech/computing/facebooks-ray-ban-smart-glasses-are-coming-what-to-expect/

Stephens-Davidowitz, S., & Pabon, A. (2017). Everybody lies: Big data, new data, and what the internet can tell us about who we really are. New York: HarperCollins.
Subway (2021). Subway [Software]. Retrieved from: ‘https://play.google.com/store/apps/details?id=com.subway.mobile.subwayapp03&hl=en_GB&gl=US’

Taleb, N. N. (2007). The black swan: The impact of the highly improbable (Vol. 2). Random house.

Taleb, N. N. (2018) Skin in the Game: The Hidden Asymmetries in Daily Life. Random House.
Taylor, D., & Maslin, J. (1970). A study of the tachistoscope. European Journal of Marketing, 4(1), 22-28.

Taylor, E., & Webster, T. J. (2011). Reducing infections through nanotechnology and nanoparticles. Int J Nanomedicine, 6, 1463-1473. doi:10.2147/IJN.S22021

Tesla (2021). Tesla [Software] Retrieved from: ‘https://play.google.com/store/apps/details?id=com.teslamotors.tesla&hl=en_GB&gl=US’

Thorp, E. (1961). A favorable strategy for twenty-one. Proceedings of the National Academy of Sciences of the United States of America, 47(1), 110.

Thorp, E. O. (2017). A Man for All Markets: Beating the Odds, from Las Vegas to Wall Street. Simon and Schuster.

Tinder (2021). Tinder [Software]. Retrieved from: ‘https://play.google.com/store/apps/details?id=com.tinder&hl=en_GB&gl=US’
van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. PloS one, 13(8), e0202442.

Vengadasalam, S., Kleinfercher, F., & Lawton, J. (2020). Bloxberg: The Trusted Research Infrastructure [White paper]. https://bloxberg.org/wp-content/uploads/2020/02/bloxberg_whitepaper_1.1.pdf

Versus (2022a). Microsoft Xbox Series X review: specs and price. Retrieved from https://versus.com/en/microsoft-xbox-series-x

Versus (2022b). Sony PlayStation vs Sony PlayStation 4 vs Sony PlayStation 5 vs Sony PS2 vs Sony PS3 20GB. Retrieved from https://versus.com/en/sony-playstation-vs-sony-playstation-4-vs-sony-playstation-5-vs-sony-ps2-vs-sony-ps3-20gb

Versus (2022c). Microsoft Lumia 950 XL. Retrieved from https://versus.com/en/microsoft-lumia-950-xl

Versus (2022d). Apple iPhone 5 Vs Microsoft Lumina 950 xl. Retrieved from

Vitalis, E., Nagy, F., Toth, Z., Forgacs, L., Bozo, A., Kardos, G., . . . Kovacs, R. (2020). Candida biofilm production is associated with higher mortality in patients with candidaemia. Mycoses, 63(4), 352-360. doi:10.1111/myc.13049

Wagner, J., Lüdtke, O., Roberts, B. W., & Trautwein, U. (2014). Who belongs to me? Social relationship and personality characteristics in the transition to young adulthood. European Journal of Personality, 28(6), 586-603.

Wang, Q., Wang, H., & Hu, F. (2018, October). Combining eeg and vr technology to assess fear of heights. In 2018 9th International Conference on Information Technology in Medicine and Education (ITME) (pp. 110-114). IEEE.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., ... & Campbell, A. T. (2014, September). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing (pp. 3-14).

Ware, S., Yue, C., Morillo, R., Lu, J., Shang, C., Bi, J., ... & Wang, B. (2020). Predicting depressive symptoms using smartphone data. Smart Health, 15, 100093.

Washburn, D. A. (2003). The games psychologists play (and the data they provide). Behavior Research Methods, Instruments, & Computers, 35(2), 185-193.

Weisel, K. K., Fuhrmann, L. M., Berking, M., Baumeister, H., Cuijpers, P., & Ebert, D. D. (2019). Standalone smartphone apps for mental health—a systematic review and meta-analysis. NPJ digital medicine, 2(1), 1-10.

Welch, N., & Krantz, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. Behavior Research Methods, Instruments, & Computers, 28(2), 192-196.

West, J. D., & Bergstrom, C. T. (2020). Calling Bullshit: The Art of Scepticism in a Data-Driven World. Penguin UK.

Whateley, L. (2018) Money - a user's guide. Harper Collins Publishers.

Whitwam R., (2019). 1 in 6 US Adults Now Own a Smartwatch [Website]. Retrieved from https://www.extremetech.com/mobile/285724-1-in-6-us-adults-now-own-a-smartwatch#:~:text=Today%2C%20about%20one%20in%20six%20US%20adults%20owns,2018%20saw%20smartwatch%20sales%20increase%20by%2061%20percent.

Wittek, P. (2014). Quantum machine learning: what quantum computing means to data mining. Academic Press.

Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In Economic analysis of the digital economy (pp. 89-118). University of Chicago Press.

Wylie, C. (2019). Mindf* ck: Inside Cambridge Analytica's plot to break the world. Profile Books.

Yaremych, H. E., & Persky, S. (2019). Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. Journal of experimental social psychology, 85, 103845.

Zimbardo, P. G., Haney, C., Banks, W. C., & Jaffe, D. (1971). *The Stanford prison experiment*. Zimbardo, Incorporated.

Zou, W., Lo, D., Kochhar, P. S., Le, X. B. D., Xia, X., Feng, Y., ... & Xu, B. (2019). Smart contract development: Challenges and opportunities. *IEEE Transactions on Software Engineering*, *47*(10), 2084-2106.

Zuckerman, G. (2019). The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution. Penguin.

Chapter II

A simple location-tracking app for psychological research

Chapter Publication status:

Geyer, K., Ellis, D. A., & Piwek, L. (2019). A simple location-tracking app for psychological research. *Behavior Research Methods, 51(6)*, 2840-2846.

Chapter I provided an overview of the disruptive potential of smartphones in psychological research. Part 1 of this thesis (Chapters II-IV) involves the development of apps.

Specific aim

1a. Develop software capable of effectively capturing the context of behaviour.

My contribution

I developed the Android app, the associated scripts for analysis and was the primary writer of the article.

According to Miller (2012) smartphones can provide insights into three areas. *Chapter II* specifically attempts to generate data to support two of these areas: context & behaviour. Logging location captures two aspects relevant about a participant: movement and surroundings. An open-source location logging android app was developed which allowed participants to continually log their location. The app utilised high-level security (the participant would be informed about what the app did & the app would only export the data to the researcher after continual compliance by the participant). Additionally, the accuracy of every data point was reported which allowed for researchers to subsequently identify time spent in specific locations with a relative degree of confidence. Finally, significant efforts were made to ensure that the data collection was as reliable as possible. Ultimately, this chapter attempts to answer the question - can an ethically-based open-source smartphone app be developed for use in psychological research?

This development went on to inform the software in *Chapters III* that can monitor other behaviours. After publication, the article attracted interest from researchers in Singapore which resulted in a further collaboration that led to a more sophisticated and customisable version of the app (Geyer, 2020; description in app store contains link to accompanying websites). Improvements involved building a app that would be more easily customisable by the researchers (like in *Chapters III*), improved security and the code changed from Java to Kotlin (embracing Android's new preferred language). A R package has specifically been constructed by other researchers to aid with the analysis of the location data generated by the App (Zipert, de Vries, & Ufkes, 2021). The article has been cited 11 times and accessed 5604 times at the time of writing.

Abstract

Location data gathered from a variety of sources is particularly valuable when it comes to understanding individuals and groups. However, much of this work relies on participants' active engagement to generate and regularly report their location. While commercial smartphone applications are available, these are often expensive and not designed with researchers in mind. In order to overcome these and other related issues, we have developed a freely available Android application, which logs location accurately, in real-time, and requires no active participation once installed. Further recommendations and R code are provided to assist with subsequent data analysis.

2.1 Introduction

Where a person spends their time can provide numerous insights into their behaviour, personality and mood (Chorely, Whitaker, Allen, 2013). For example, location measures derived from a smartphone can be predictive of depressive symptoms and levels of social anxiety (Huang, et al., 2016; Palmius et al., 2017). Other research has also shown that individuals with comparable personalities report accessing similar locations (Noe, Whitaker, Chorley, Pollet, 2016). While these studies remain important, critics have argued that comparatively little research has actually been conducted regarding what drives peoples' movements, and what is psychologically important about the locations that people choose to occupy (e.g., Rauthmann et al., 2014). While smartphones provide huge potential in this regard with almost every device containing a GPS sensor, there remains a lack of suitable software that is freely available for those working within psychology and the social sciences more generally (Piwek & Ellis, 2016). Researchers struggle find appropriate alternatives from commercial application repositories, such as Google play store (Google, 2017a) or the app store (Apple, 2017). This is largely because these applications have not been developed with social research in mind. Many commercial applications, for example, often struggle to strike a suitable balance between high levels of accuracy and duration of logging, which are methodologically important for location based research (Palmius et al., 2017).

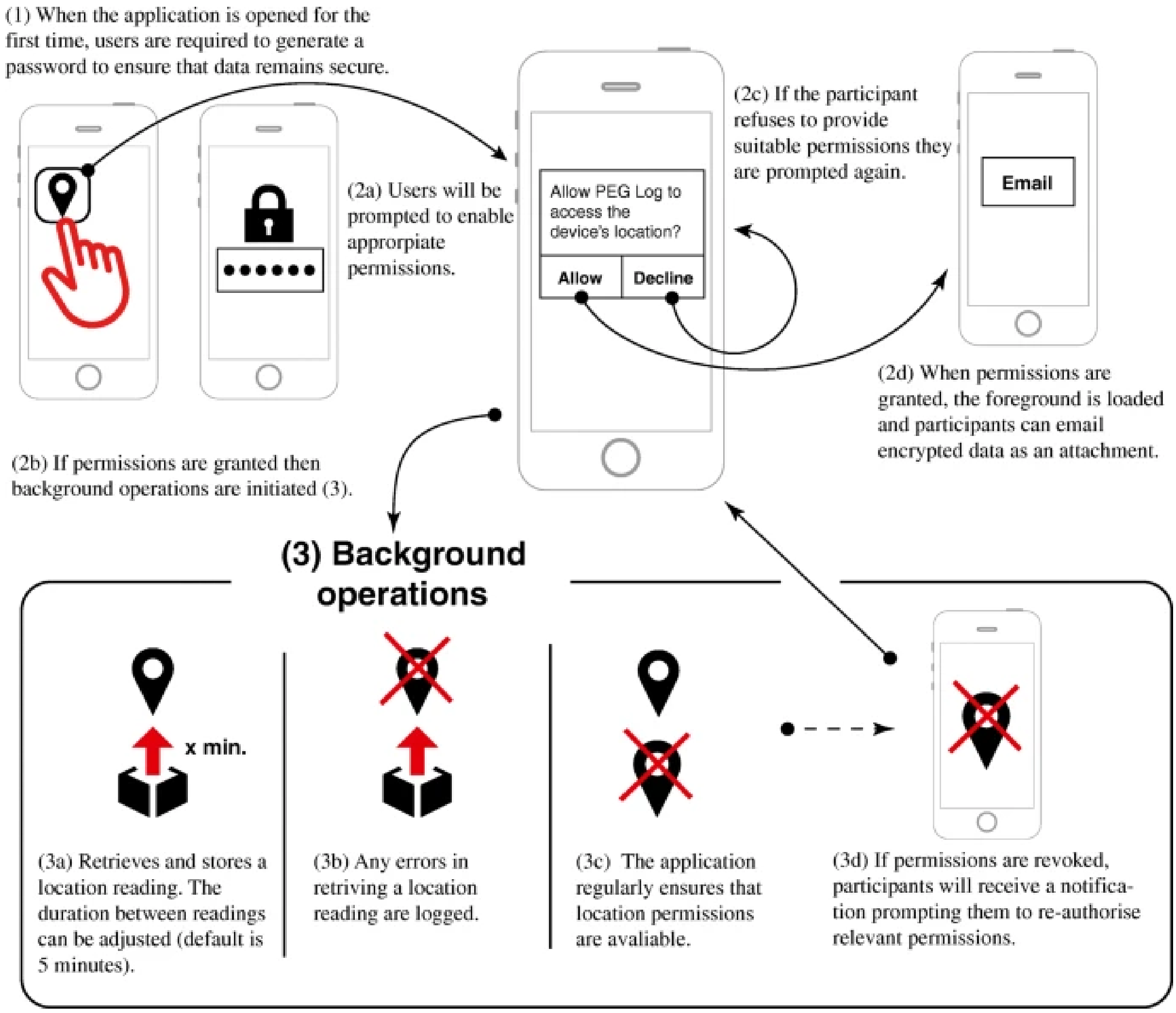
Researchers unable to access location data via smartphone applications have instead relied on other innovative methods. For example, location databases harvested from social media websites have demonstrated that it is possible to predict personality from where a person chooses to spend their time (Chorley, Whitaker, Allen, 2015; Noe, Whitaker, Chorley, Pollet, 2016; Song, & Lee, 2015). However, this method presents new limitations because using social media to sample multiple locations is likely to only include the reporting of socially desirable locations (Schwartz, & Halegoua, 2015). This effect may be magnified further as social media users are motivated to selectively report their location in order to maintain or boost their social status (Fitzpatric, Birnholtz, & Gergle, 2016; Guha, & Birnholtz, 2013; Schwartz, & Halegoua, 2015). Other correlational designs have relied on self-report via experience sampling smartphone applications (e.g., Sandstrom, Lathia, Mascolo, Rentfrow, 2017). Like social media capture, the reporting of every location that an individual visits requires an extensive amount of effort. Further, the accuracy and resolution of the data is often poor. As a result, data generated from either method provides a patchy account of where a person spends their time.

2.1.1 Implementation

In order to overcome previous methodological limitations, we have developed a freely available application (*PEG LOG*) that records the location of an Android smartphone. This is an attempt to enhance the quality and quantity of data that is available to researchers studying the significance of individual and group movements. Additionally, we wish to prompt transparency and replication by making the source code and supplementary materials freely available. The application was built to require minimal effort from participants while remaining more informative than any self-reported measure.

Summary of application architecture

The application runs on Android devices and available from Google Play store (https://play.google.com/store/apps/details?id=com.location.android.peglog&hl=en). It was designed in order to provide regular updates while acknowledging the limitations of location measures. For example, GPS signals are typically inaccessible from inside a building, but the application can switch to rely on other available sources that report location, e.g., Wi-Fi and Network signals. However, it should be noted that both these signals remain less accurate than GPS (Canzian, & Musolesi, 2015).



[Figure 2.1: demonstrating the foreground and background operations of the application. The only aspect of the application accessible by participants is the main activity page, which requests the relevant permissions (location and external files), and allows participants to delete and/or email files. After the application is opened for the first time (1) a user will be prompted to give location permission. If the participant refuses to provide permission they are prompted again (2). If location permission is granted then background operations (3) are initiated. Background operations include: storing location data if other applications are accessing within 15 seconds of the last location update (4), retrieving a location reading after 60 seconds (5), and checking every five minutes when location permissions remain granted. If the permissions are no longer enabled, the participant will receive a notification (7) prompting them to re-enable location permissions. Notifications will lead participants to a screen where they can re-enable the permissions, which will re-start background services (8) when enabled. After location permission is granted, the foreground is loaded (9) and participants can email raw data as an attachment (10) or delete the stored file (11).]

Installation

The installation process was designed to be straight-forward and requires almost no time or commitment from participants. In order for data collection to begin, participants are required to access the application on the Google Play store. Following installation, they will then be required to open the application. An additional step is required if the participant possess an Android smartphone with an API of 23 or higher. These participants will be prompted to give explicit permission that the application can access location services. Once the application is installed participants simply have to open the application and confirm that data collection can commence. It is advisable that all participants send some pilot data to the researcher at the beginning of the study to ensure location tracking is underway as expected.

2.2 Foreground operations

2.2.1 Data storage and export

Data is stored locally and in a manner that ensures that only the specific application can access this information. If the participant chooses to withdraw from the study they simply have to uninstall the application and data collection will cease, however, data already collected will remain on the smartphone. To export data, participants can select ‘email’ and then enable permission for the application to write in external memory (Figure 1). They can then use any email account to send the file with their data stored in an attachment. The file returned is a lengthy string, which comprises of latitude, longitude, accuracy (the size of the radius if researchers are to have 68% confidence that the smartphone is inside the area) and a timestamp (UNIX time – amount of milliseconds to elapse since midnight 1st January 1970; see Supplementary Materials for an example of raw data). This will be labelled at the top of each data file and will repeat each time the application is restarted. If the phone is restarted this will also be recorded on the data file. Very little processing of the location data is carried out within the application itself order to maximise battery life. To illustrate some simple data processing and analysis operations, the included R-script can be used to process raw data and generate basic visualisations. For example, heat maps can illustrate where a person has spent the majority of their time (see Supplementary Materials).

2.2.2 Deleting files

If the participant forgets about the study, or does not wish to proactively comply with the requirements of the study then researchers cannot retrieve these data. This was a conscious decision in to order to comply with standard ethical guidelines.

However, following data collection, a participant can then delete files from their device. This is done by clicking the delete button and removes data from both external and internal memory. A password is required to complete this operation (“oeg”; all lowercase). This function is password protected to ensure that accidental deletion is not possible. The result of this operation will be confirmed via a textbox (Figure 1).

2.3 Background operations

2.3.1 Accessing location data

The application relies on the FuseLocationProvider (Google, 2017b). This provides access to GPS, Wi-Fi, and Network analysis in order to retrieve latitude, longitude, accuracy in meters with regards to the radius of confidence and a UNIX timestamp. The application is considered high priority meaning that the most accurate reading available is provided regardless of battery expenditure. The order of favourability of trace (in relation to accuracy) is therefore: GPS, Wi-Fi followed by network analysis (Canzian, & Musolesi, 2015). A location update is requested by default every 60 seconds. If another application requests a location reading from the smartphone after 15 seconds of PEG LOG receiving a location update then the PEG LOG will also store this information. It is worth noting that retrieving data from other applications that request location data improves battery life. However, other applications may not have access to an equivalent resolution of location data and this may reduce overall accuracy. Similarly, PEG LOG can be combined with a variety of other applications or methodologies that extract information from a smartphone (e.g., experience sampling applications) and continue to run in the background.

Customisation

Which location data source (GPS, Wi-Fi, etc.) is used by default, and the frequency of location updates can be customized by following a simple modification to the original source code. This is outlined within one non-expert friendly file: Constants (this file explaining the project structure is available via the associated GitHub address). Following customisation, the application can then be redistributed on the Google Play store.

2.3.2 Resilience of the application

We have identified five potential ways that the application could be prevented from functioning. Participants could inadvertently stop data collection by: (1) closing the application, (2) clearing tasks running in the foreground, (3) turning off the phone, (4) forcing closure of all running applications, or (5) uninstalling the application. Addressing these issues in order, if the foreground section of the application is closed then the background service will continue to run or restart if it has also been closed. If all foreground applications are closed, then the background service will automatically restart. If the phone is turned off, upon restarting, the application will automatically resume and continue running as a background operation. This will, in turn, be documented in internal memory and mark an interruption of data collection due to a restart event. However, if a force closure of all applications occurs then the participant will be required to open the application again in order to continue with data collection. Uninstallation is interpreted as a desire to withdraw from the study and uninstalling the application has been maintained will stop data collection. Finally, if a participant does not have location permissions enabled for a period of 5 minutes the application will send the participant a notification. This will notify them that the location permissions should be enabled. This also informs the participant that if they click on the notification, it will take them to relevant settings where they can enable location permissions.

2.4 Results & Discussion

A complete review concerning how location data can be analysed is beyond the scope of this paper. However, broadly, there are two key ways or analysing real-time location data. Firstly, location points can be placed into topologies such as: café, library, nightclub, etc. Locations via this method can be further characterised based on how they relate to other geographic databases, e.g., census records, crime statistics, foursquare database (Canzian, &

Musolesi, 2015; Chorley, Whitaker, & Allen, 2014; Rauthmann, et al., 2014). Secondly, the movements a person engages can be characterised in a number of ways (Canzian, & Musolesi, 2015). This can include information relating to: distance travelled, radius of gyration, etc. For example, recent psychological research has shown that an analysis, which includes information relating to both journey and destination, is incrementally more valuable (Huang, et al., 2016). This combines both approaches outlined above, however, there remains potential for these simple analyses to develop further as real-time location data continues to become easier to collect. We have therefore provided additional supplementary R-code to assist with these developments. This marked-up code will process raw location data, prepare data for analysis, and generate some basic visualisations (Figure 2).

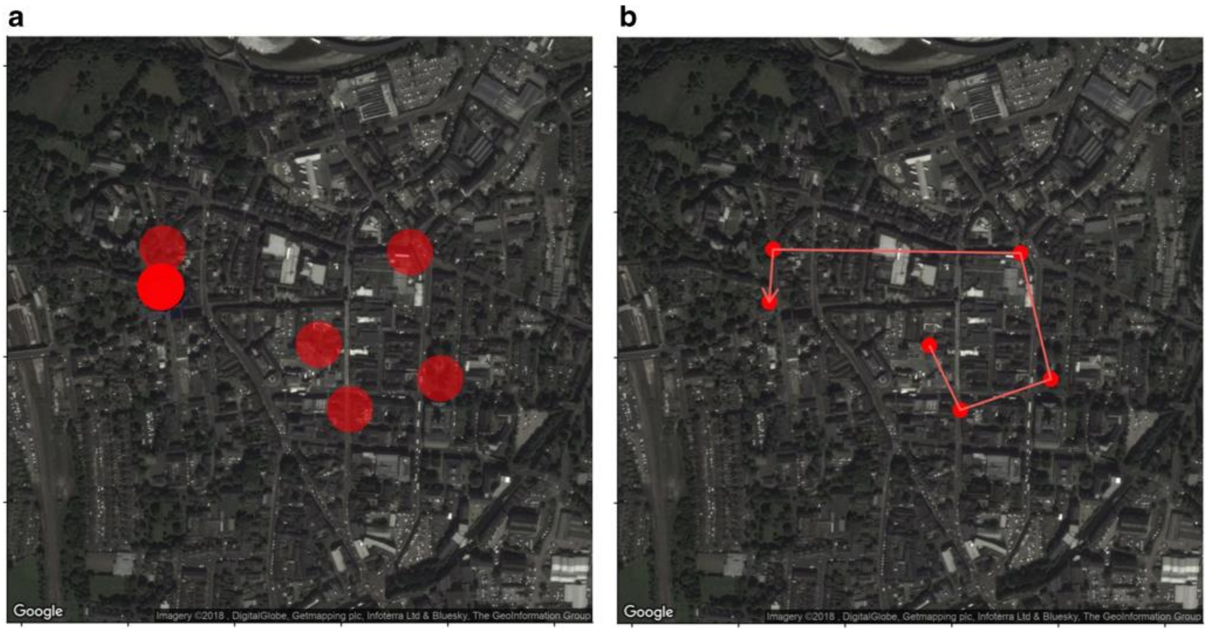


Figure 2.2: A simple visualisation from a single day of location tracking. This includes a heat map (A) to illustrate where a participants spent the majority of their time and (B) a path map were movement has been plotted over the course of 24 hours

2.5 Conclusion

Researchers who have collected location data from smartphones and other digital devices have previously found this digital trace to be both predicative of future movements and a variety of other individual differences (Chorely, Whitaker, Allen, 2013). However, research has often drawn conclusions based on incomplete recordings of location and these remain problematic. Overcoming these limitations for social science remains important in order to pre-empt the well-documented issues with self-reported data, especially when recording location information in real-time. In summary, we have presented a freely available location tracking application and analysis code, which will allow many researchers across a variety of disciplines to conduct rigorous research into individual and group movements.

Declarations

Availability of data and material

The application is freely available at this website:

https://play.google.com/store/apps/details?id=com.location.android.peglog&hl=en

Source code is available from the following GitHub: https://github.com/kris-geyer/PEGlog

Data and supplementary materials are attached as part of this submission

Competing interests

The authors report no conflicts of interest.

Funding

This work was part funded by the Centre for Research and Evidence on Security Threats [ESRC Award: ES/N009614/1].

Authors' contributions

KG developed and tested the application and wrote the first draft of the manuscript. DAE contributed to the writing of the manuscript and supplementary materials. LP contributed to the writing of the manuscript.

References

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: tools to compare estimated and real-world smartphone use. *PloS one*, *10*(10), e0139004.

Android (2017, August 22) *Background Location Limits*. Retrieved from: https://developer.android.com/about/versions/oreo/background-location-limits.html

Apple (2017) *App store*. Retrieved from URL: https://www.appstore.com/

Canzian, L., & Musolesi, M. (2015, September). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 1293-1304). ACM.

Chorley, M. J., Whitaker, R. M., & Allen, S. M. (2015). Personality and location-based social networks. *Computers in Human Behavior*, *46*, 45-56.

Fitzpatrick, C., Birnholtz, J. P., & Gergle, D. (2016, September). People, places, and perceptions: effects of location check-in awareness on impressions of strangers. In *MobileHCI* (pp. 295-305).

Geyer, K. (2020) Peg Log 2.0 [software]. Retrieved from https://play.google.com/store/apps/details?id=kg.own.kotlin&hl=en&gl=US

Google [Computer software] (2017b) *FusedLocationProviderAPI* Retrieved from URL: https://developers.google.com/android/reference/com/google/android/gms/location/FusedLocationProviderApi

Guha, S., & Birnholtz, J. (2013, August). Can you see me now?: location, visibility and the management of impressions on foursquare. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services* (pp. 183-192). ACM.

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83-90.

Huang, Y., Xiong, H., Leach, K., Zhang, Y., Chow, P., Fua, K., ... & Barnes, L. E. (2016, September). Assessing social anxiety using GPS trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 898-903). ACM.

Lopez-Fernandez, O. (2017). Short version of the Smartphone Addiction Scale adapted to Spanish and French: Towards a cross-cultural research in problematic mobile phone use. *Addictive behaviors*, *64*, 275-280.

Miller, G. (2012). The smartphone psychology manifesto. Perspectives on psychological science, 7(3), 221-237.Google (2017a) *Play store*. Retrieved from URL: https://play.google.com/store?hl=en

Noë, N., Whitaker, R. M., Chorley, M. J., & Pollet, T. V. (2016). Birds of a feather locate together? Foursquare checkins and personality homophily. *Computers in Human Behavior*, *58*, 343-353.

Palmius, N., Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Geddes, J. R., Goodwin, G. M., & De Vos, M. (2017). Detecting bipolar depression from geographic location data. *IEEE Transactions on Biomedical Engineering*, *64*(8), 1761-1771.

Piwek, L., & Ellis, D. A. (2016). Can programming frameworks bring smartphones into the mainstream of psychological science?. *Frontiers in psychology*, *7*.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., ... & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677.

Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P., & Mohr, D. C. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ*, *4*, e2537.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research*, *17*(7).

Sandham, A., Ormerod, T., Dando, C., Bull, R., Jackson, M., & Goulding, J. (2011). Scent trails: Countering terrorism through informed surveillance. *Engineering Psychology and Cognitive Ergonomics*, 452-460.

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality*, *69*, 96-101.

Schwartz, R., & Halegoua, G. R. (2015). The spatial self: Location-based identity performance on social media. *New media & society*, *17*(10), 1643-1660.

Snap Inc (2017) *Snapchat* (Version 10.17.5.0) [Mobile application software]. Retrieved from: https://play.google.com/store/

Solon, O (2017, 23 June) Snapchat's new map feature raises fears of stalking and bullying. *The guardian*, Retrieved from: https://www.theguardian.com/technology/2017/jun/23/snapchat-maps-privacy-safety-concerns

Song, H. Y., & Lee, E. B. (2015). An analysis of the relationship between human personality and favored location. *AFIN 2015*, 12.

Ziepert, B., de Vries, P. W., & Ufkes, E. (2021). “Psyosphere”: A GPS Data-Analysing Tool for the Behavioural Sciences. Frontiers in psychology, 12, 1650.

Chapter III

Open source smartphone apps and tools for measuring, quantifying, and visualizing technology use

Chapter Publication status: Published

Geyer, K., Ellis, D. A., Shaw, H., & Davidson, B. I. (2021). Open source smartphone apps and tools for measuring, quantifying and visualizing screen time. *Behaviour Research Methods*. 54 (1) 1 – 12

Specific aim

1b. Develop software capable that captures behaviour.

My contribution

I developed the Android app, the associated scripts for analysis, created the validation app, associated websites for the customization of the app and was the primary writer of the article.

Chapter II reported on the development of a smartphone apps which reliably logged the location of an Android smartphone and monitored neurological activity. From our efforts with PEG log much was learnt, and the poor design decisions could be avoided. For instance, customisation was done via scanning of a QR code making the app more responsive to the needs of the researchers. Again, better design decisions were identified through the development of past apps. For example, there were issues associated with passively recording behaviour. When the phone is inactive the continual logging of behaviour is sometimes restricted. However, for monitoring other types of digital behaviour this may not necessarily be affected by this problem. Developing a psych app that monitors screen usage would be an effective exploration of what smartphones allow developers to monitor conveniently, is virtual behaviour easily monitored comparative to non-virtual behaviour? Additionally, a tool to accurately record smartphone behaviour would be very useful for researchers interested in smartphone 'addiction'.

In this chapter, I designed another android psych app that recorded smartphone usage. Therefore, I developed an app which can both passively monitor smartphone usage and also query a database that holds high resolution details about how the smartphone was previously used. This chapter also attempts to understand if apps that monitor digital behaviour require a different design to other passive sensor designs (e.g., location).

Ultimately, this app becomes core to later research documented within Part 2 of the thesis because it can be customised to fit multiple research requirements. It removes the involvement of demand-characteristics in experiments by gathering smartphone usage data collected up to five days prior to the experiment starting. After the recent publication, this app has been cited 4 times and downloaded 268 times, at time of writing. The capabilities of this work has prompted much collaboration with other researchers internationally including Japan, Nigeria and Spain.

Open source smartphone app and tools for measuring, quantifying, and visualizing technology use

Kristoffer Geyer[1], David A. Ellis[2], Heather Shaw[1], and Brittany I. Davidson[2,3]

[1]Department of Psychology, Lancaster University, UK

[2]School of Management, University of Bath, UK

[3]Department of Engineering, University of Bristol, UK

Corresponding author: David Ellis, School of Management, University of Bath, Bath, BA2 7AY, United Kingdom. E-mail: dae30@bath.ac.uk

Abstract

Psychological science has spent many years attempting to understand the impact of new technology on people and society. However, the frequent use of self-report methods to quantify patterns of usage struggle to capture subtle nuances of human-computer interaction. This has become particularly problematic for devices like smartphones that are used frequently and for a variety of purposes. While commercial apps can provide an element of objectivity, these are ‘closed’ and cannot be adapted to deliver a researcher-focused ‘open’ platform that allows for straightforward replication. Therefore, we have developed an Android app, which provides accurate, highly detailed, and customisable accounts of smartphone usage without compromising participants privacy. Further recommendations and code are provided to assist with data analysis. All source code, materials and data are freely available (see links in supplementary materials section).

4.1 Introduction

Human-computer interactions have become ubiquitous and continue to shape individuals and society (Ellis, 2019). For instance, many people in the general population will interact with their smartphone over 80 times a day in order access a variety of online services (Andrews, Ellis, Shaw, & Piwek, 2015; Ellis, Davidson, Shaw, & Geyer, 2019). As a result, the way in which individuals and groups use these technologies has provided new opportunities for research. Psychological science has specifically focused much of its efforts on understanding how technology use may impact our health, social processes, and cognitive functioning (Ellis, 2020). For example, general smartphone use has been associated with a variety of negative outcomes including depression (Elhai, Dvorak, Levine, & Hall, 2017), anxiety (Richardson, Hussain, & Griffiths, 2018), disrupted sleep (Rosen, Carrier, Miller, Rokkum, & Ruiz, 2016), cognitive impairment (Clayton, Leshner, & Almond, 2015), and poor academic performance (Lepp, Barkley, & Karpinski, 2015). This repeats a pattern of research priorities, which previously focused on the negative impacts of many other screen-based technologies, systematically moving from television and video games, to the internet, and social media (Rosen et al., 2014; Davidson & Ellis, 2019). In contrast, pre-registered studies have suggested that technologies, which were previously deemed problematic including social media and video games have negligible or positive associations with well-being (Orben & Przybylski, 2019a; Johannes, Vuorre & Przybylski, 2020).

The majority of research in this area shares a similar methodology when capturing technology usage. This typically relies on asking participants for a duration estimate or a qualitative reflection concerning their own experience rather than objectively measuring behavior from a device (Ellis, 2019). For example, requests for single duration estimates might ask: ‘how much time do you spend on your smartphone per day?’. Psychometric scales that are also common and include a range of items that attempt to quantify ‘problematic’ or ‘addictive’ patterns of usage (Ellis, 2019). While such measurements are typical across social psychology, they have well established limitations when attempting to describe behaviors or understand related psychological impacts (Baumeister, Vohs, & Funder, 2007; Doliński, 2018; Hinds & Joinson, 2019; Sassenberg, & Ditrich, 2019). Single duration estimates are unable to capture the range of experiences provided by modern technology and survey instruments poorly correlate with a variety of objectively measured behaviors (Boas, & Ling, 2013; Ellis et al., 2019). For example, picking up a smartphone is habitual and often occurs unconsciously throughout the day making it difficult to self-report accurately (Andrews, Ellis, Shaw, & Piwek, 2015; Ellis, Davidson, Shaw, & Geyer, 2019; Ellis, 2019).

Researchers have suggested that the automated tracking of technology related behaviors are valuable, but remain technically challenging to collect (Orben & Przybylski, 2019b, Johannes, et al., 2019). However, a number of commercial apps can now quantify high-level aspects of smartphone usage including total daily usage and number of pick-ups (Ellis et al., 2019). These include pre-installed tools: digital health (Google LLC, 2019) for Android systems, and Apple’ Screen Time for iOS devices (Apple, 2019). While these apps can provide a more objective account of a person’s smartphone usage, they have several limitations. First, data is of a low resolution and only provides daily metrics of usage. In order to answer specific research questions, hourly or minute-by-minute metrics are essential however, the majority of third-party apps (e.g., *Moment, App Usage*) are, at the time of writing, unable to report the distribution of smartphone use durations across multiple 24-hour periods. Second, commercial apps cannot be re-configured to provide additional functionality. For example, there is often no way to export raw data, requiring participants to manually transfer statistics into survey responses (e.g., Ellis et al., 2019). As a consequence, commercial apps are ‘closed’ to the extent that researchers are unable to access source code, making it difficult to assess the reliability and validity of data collected.

In an attempt to circumvent some of these limitations, researchers have developed programming frameworks (e.g., *Funf in a Box* and *Aware*), which can facilitate the development of specific apps that could record technology-related behaviours (Aharony et al., 2011; Ferreira, Kostakos & Dey, 2015). However, these frameworks were predominantly designed to capture data from a variety of smartphone sensors. While the extensive APIs and associated libraries provide many data collection possibilities, this will always require significant development and customisation before it becomes useful for researchers and safe for participants (Piwek, Ellis & Andrews, 2016). For example, a server will often be required to receive data remotely and researchers must implement sophisticated network protocols to maintain the security of data during transfer. Creating research-orientated applications from these frameworks therefore remains challenging for researchers who are unfamiliar with mobile app development and who wish to ensure data succinctness.

To mitigate these issues and allow the research base to engage with objective methods, we present a customisable Android smartphone app – Usage Logger; and its associated R scripts, python notebooks (Jupyter), and websites (see supplementary materials). Together, these resources provide researchers with a succinct way to capture a variety of smartphone usage behaviors. This includes the quantification of time spent on a device, specific app use and notification activities.

Usage Logger timestamps every interaction the user has with their phone, which can generate a sophisticated understanding of general and specific technology usage. In addition, it can also extract historical data from the device, which addresses concerns of social desirability. As a result, the tools described here will help contribute to existing and new avenues of research. Specific research designs might consider, for example, if there are points in time where a participant's usage is habitual or more entropic (Aledavood, Lehmann & Saramäki, 2018) or if usage was prompted by a notification or goal-directed. Given long-standing concerns regarding the impact of new technology, these resources will also allow researchers to better understand if certain patterns of usage have disproportionate associations with different psychological process and outcomes (Ellis, 2019). The rest of this article provides a comprehensive overview of the app and details how researchers can customize its operation, understand the data collected and generate (or replicate) usage variables. All analysis scripts and associated software are freely available.

4.2 Summary of Architecture

The overall aim was to develop an app and supporting resources that will allow psychologists and others within the social sciences to conduct research that involves measuring smartphone technology interactions. The first step in the development of such a tool was to define the basic criteria that it needed to fulfil. For the aims of this project, these resources should: (a) provide open source code so resources can be scrutinised and/or developed by other researchers.; (b) record a variety of technology interactions, while ensuring data succinctness (i.e., only data required is collected); (c) remain intuitive, practical, quick, and easy to use for groups of researchers who vary considerably in their computational literacy and finally (d) ensure data remains secure and protect participant's privacy with the opportunity to withdraw during any study (data remains on a device until participants wish to share it with researchers).

A variety of models have been proposed concerning software development lifecycles (Van Vliet, 2008). During development, we predominantly relied on a prototyping model because the system was developed alongside end users (researchers and participants) to improve each iteration of the software. As with related developments, Android was chosen as the initial development OS, as it offered technical and methodological flexibility at the best cost–performance ratio (Geyer, Ellis and Piwek, 2019; Keil, Koschate & Levine, 2020). 80% of the worldwide market also run Android related software (Keil, Koschate & Levine, 2020).

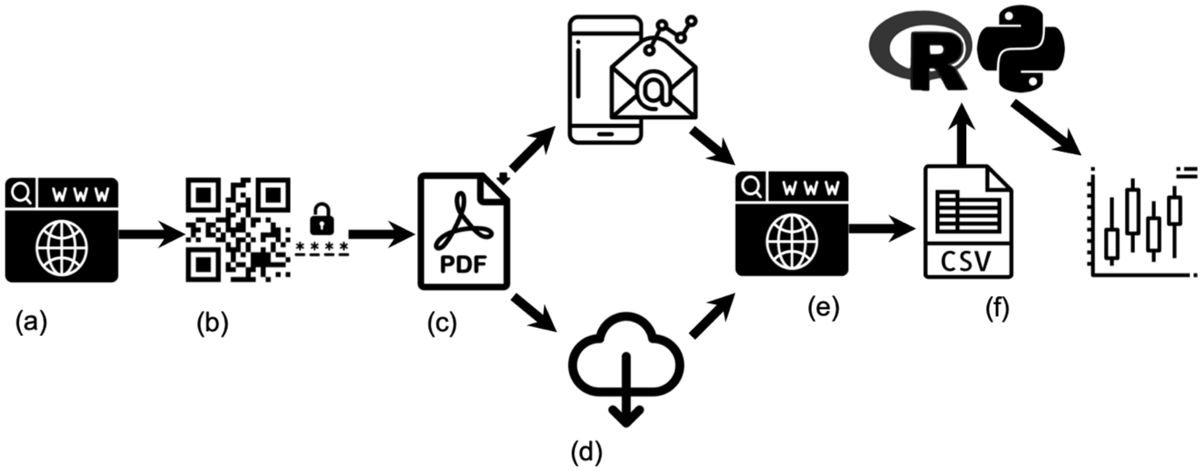
The system consists of four major elements: a website to customise what data the smartphone app collects and/or retrieves, a Usage Logger app that enables data collection, a second website to assist with data processing, and a series of R scripts and python notebooks (Jupyter) that contain live code and visualisations to help with analysis (Figure 1). In line with our first aim (a) concerning open science, all 3rd party open-source libraries, which are essential to the functioning of the presented resources are freely available (Table 1). As a result, researchers can take any single element, or combine materials as they wish and customize them as necessary. In the following sections, we describe each in detail.

Library name	Element	Link	Version	Function
JQuery	Customisation website	https://blog.jqueryui.com/2015/03/jquery-ui-1-11-4/	1.11.4	Succinct javascript
Canvas2Image	Customisation website	https://github.com/hongru/canvas2image	NA	Converts QR code to png image
Qrcode	Customisation website	https://davidshimjs.github.io/qrcodejs/	NA	Generates QR codes
Timber	App	https://github.com/JakeWharton/timber	4.7.1	Facilitates communication between app and developer
Dm77/barcodescanner	App	https://github.com/dm77/barcodescanner	1.9.13	Scans QR codes
Code Scanner	App	https://github.com/yuriy-budiyev/code-scanner	2.1.0	Scans QR codes post Android 8.1
Gson	App	https://github.com/google/gson	2.8.2	Convert Java objects to JSON
Armadillo Encrypted Shared Preferences	App	https://mvnrepository.com/artifact/at.favre.lib/armadillo/0.9.0	0.9.0	Encrypts data
SQLcipher	App	https://github.com/sqlcipher/android-database-sqlcipher	4.0.0	Constructs encrypted SQL databases
Jabit Spongy Cryptography	App	https://mvnrepository.com/artifact/ch.dissem.jabit/jabit-cryptography-spongy	2.0.4	Facilitates cryptographic calculations

iText	App	https://github.com/itext/itextpdf/releases	5.5.10	Constructs encrypted pdfs
PDF.js	Data processing website	https://mozilla.github.io/pdf.js/	NA	Interacts with PDF via javascript

[Table 4.1. A list of 3rd party libraries used by Usage Logger and associated websites.]

[Figure 4.1. Overview of Usage Logger: (a) specification of configuration capabilities online; (b) a QR code is generated by the website and the app generates a secure password to encrypt all information and commences data collection; (c) post-data collection, PDFs are generated; (d) these files can be exported via email, to another app or cloud service; (e) files can be decrypted via a second website; (f) this generates a .csv file, which can be processed using the provided R scripts and python notebooks. All software and materials are open source and freely available in line with recommendations outlined by the UK Reproducibility network (Turner et al., 2019)]



4.3 Data Sources and Customisation

A researcher will first have to specify via the customization website (see supplementary materials for web addresses) what data will be collected by the app. In line with our second aim (b), this ensures that researchers are only collecting data that is essential to their research question where ethical approval has been received by a relevant committee. The customization website allows for three different data sources to be collected: *contextual*, *past usage*, and *continuous* usage. We outline these in more detail below. These summaries are also available as part of our companion website (see supplementary materials).

Contextual data sources provide information regarding the context in which apps are used. There are three sources of data that can be collected: installed apps, permissions requested, and response to permission requests. Installed apps have previously been found to offer insights regarding personality (Xu, Frey, Fleisch, & Ilic, 2016). However, permissions requested and granted may also provide insights about a person's attitudes towards privacy (Reinfelder, Schankin, Russ, & Benenson, 2018). For example, a user with Facebook Messenger (Facebook, 2019) installed will be prompted to provide permissions that allow the app to access location sensing capabilities. This allows Messenger to send location updates to their friends and contacts. Responses to such requests also allows researchers to understand what data an associated app actually has access to via the smartphones operating system. Using the contextual capability, the app generates a file containing this data saved as "*context.pdf*".

Continuous logging collects information about multiple smartphone interactions as they occur **after** installation of Usage Logger. Researchers can decide what types of data should be collected including: when the phone is on/off, what apps are opened, if the apps list is updated (though installation or uninstallation) and when apps send notifications. The app will by default always document if a smartphone is restarted.

At a basic level, researchers are provided with information about when a participant is using their smartphone. This can be extended to include usage as it relates to specific apps and when these are installed or removed. Analysis of notifications can help researchers differentiate between smartphone interactions that have been directed by an individual versus notifications that drive usage. All events captured by the app are UNIX timestamped and placed in a file called: "*continuous.pdf*".

Past usage is based on how a participant has used their device previously, **before** installation of Usage Logger. Akin to an internet browser history, this data is stored and maintained by Android devices. Usage Logger in this instance queries the appropriate database. This data can be accessed after a participant approves a specific permission – *package usage stats* and provides information concerning usage events and statistics. Usage events document when a person had turned their screen on or off, switched app, switched their phone on/off, and details on how the operating system managed apps (Android, 2019a). Our testing as part of previous research (e.g., Shaw, Ellis, Geyer, Davidson, Ziegler and Smith, 2020) and reported validation suggests that this history typically extends to a period of seven days. This might be useful for a variety of studies as previous research suggests that only 5 days of retrospective data is sufficient to be representative of a person's overall smartphone use (Wilcockson, Ellis, & Shaw, 2018). It also avoids social desirability effects because unlike continuous logging, a participant cannot change their behavior before data is retrieved. Usage Logger returns a UNIX timestamp of when an event occurred, which app was involved, and what type of event occurred. This data is supplied in a file called "*usage.pdf*".

4.3.1 Order of Collection

The customization website allows researchers to modify what data is collected by selecting specific sources. The order may be determined by dragging and ordering these sources accordingly. These decisions are likely to be driven by specific research question. For instance, if a researcher wishes to review the impact of participants having data collected continuously then they might collect five days of past usage and then contrast this with continuous data where a participant is aware that their usage is being recorded. Otherwise, if a researcher wants a higher degree of certainty regarding the collection of suitable data, they can collect up to seven days of data prospectively and retrospectively query the same period afterwards. This will ensure that if prospective data logging was paused at any point in time, higher-level missing data should be available via a retrospective trace as it relies on a different method of collection. It should be noted that participants have to trigger the switch from one data source to another (e.g. continuous logging to past usage). This feature was added to ensure that participants remain in complete control of the data collected from their device (see consent and data security).

4.4. Installation and Operation

In line with our third aim (c), Usage Logger aims to be straightforward to use by researchers and participants. Following installation by participants, which can be accomplished via the Google Play Store (see supplementary materials), the app (~10.6 MB in size) will request permission to access the camera so it can scan QR codes. After researchers specify their requirements (via tick boxes), QR codes are generated by the customisation website. These QR codes contain all the details Usage Logger needs to configure itself and perform the desired data collection. This method of configuration was selected because it allows researchers to request that participants download an identical app, which can be customized quickly and accurately without having to rely on further input from end users. This also removes the need to modify source code and, in turn, reduces the possibility of programming errors. QR codes also provide additional flexibility for researchers as they can be quickly made available to participants as part of physical research materials or placed online.

Once a QR code has been scanned, participants will encounter four dialog boxes containing information on: the purpose of the app, the type of data being collected, and security/data protection measures. These messages also invite participants to review the app's privacy policy (see supplementary materials). A password is then generated and participants are asked to approve several additional permission requests (dependent on data sources collected). These include, usage access, which allows the app to query a database that is maintained centrally by android devices about what apps were previously used. In addition, notification access allows the app to detect notifications. After suitable permissions are provided, the app will begin data collection in line with what was assigned during customisation. Finally, a number of background processes that collect data will begin.

4.5 Data Security and Consent

Previous work by the authors has focused on the transmission of highly sensitive location data (Geyer, Ellis, & Piwek, 2019). Following suit, the security and safety of participant's data again remains paramount (aim (d)). Usage Logger has been developed to ensure that participants have control over their data while it is collected. However, we would not recommend running Usage Logger on any device that has been rooted because this may undermine data security protocols

In the first instance, Usage Logger generates a password to protect data files. Relying on participants to generate passwords to protect their own data is notoriously difficult (Nelson & Vu, 2010) as these are often vulnerable to cracking. Hence, we elected to generate passwords automatically. Our solution was to utilise ‘user-generated-randomness’ (Alimommeni, 2014), which is the insignificant elements of participants actions that can be employed to seed a random generator. In usage logger, the app uses a UNIX timestamp, which is generated from each of the four times a participant confirms that they have read a message about how the app functions. These values are then stored in a random order. A value is randomly queried from this list and used to seed a random character generator to create a password. The random nature of these passwords makes them less vulnerable to dictionary attacks, which rely on databases of previously leaked passwords (Bellovin & Merritt, 1992). A variety of characters also make the password more resistant to brute-force attacks (Pliam, 2000). Participants do not have to remember this password, but can recall it from the app at any time. Of course, this layer of security relies on the Android device itself remaining secure and participants not sharing their password. We therefore recommend that participants have a screen-lock or related system enabled on their device to prevent unauthorised access.

Data collected by Usage Logger is then stored in a SQLcipher database (Android, 2019b), and only Usage Logger can access the database provided normal security protocols are not extensively compromised. To protect users further, data is protected by The Advanced Encryption Standards (AES) 256-bit block cipher in a SQLcipher database (Zetetic, 2019; 14 rounds of substitution and permutation utilized in order to encrypt the data). Data also remains secure while being exported by being inserted into an AES 256-bit encrypted pdf-file after being extracted from the encrypted database. The pdf is then supplied to the app (Android, 2019c), which is capable of exporting data over an email or an alternative method.

The decision to develop an app with email capabilities to export the files ensures that researchers do not need to set up a cloud-based storage system. Our source code could of course be modified or incorporated into any cloud-based development in the future. In the provided system however, participants can straightforwardly remain in control of their data during collection. Presently, participants can withdraw before providing any data to researchers. In order for participants to pass their data onto the research team they must; install the app, read the instructions on how the app works, approve permissions, allow time to pass while data collection occurs, manually export their data, and provide their password to a researcher.

This password handover process aims to strike a balance between providing functionality (so researchers can actually use the tools) and security (so data is remains safe). It also has to be considered alongside how damaging a data breach might be for the individual. There are therefore several options when securely transferring a password from participant to researcher, which are ordered from most to least secure. First, participants could simply read out their password to a researcher in a secure laboratory environment. Second, peer to peer encryption could be utilised using Telegram or similar apps, which sit outside an email ecosystem, to transmit passwords (Barthelmäs, Killinger & Keller, 2020). Finally, participants could send their password in a separate email that does not include raw data.

At any point in this process, a participant can uninstall Usage Logger and all data will be erased. Researchers should request that participants uninstall the app after emailing data unless they wish to collect more data for personal use as continuous logging, if enabled, will resume. Beyond this, the user experience of the app has remained minimalistic to discourage excessive interaction with the app and reduce the likelihood of demand characteristics impacting behavior. However, to ensure that participants are always aware that their behavior is being tracked during continuous logging, a notification will indicate data collection is ongoing. This notification also allows background data collection to run reliably (Geyer, Ellis and Piwek, 2020).

4.6 Reliability

Usage Logger has been designed to sustain continuous logging for substantial periods of time. The amount of data that can be logged will be limited to the free space available, however the storage capacity of most modern smartphones is unlikely to impose any limits on how much data could be collected. However, some situations or actions will naturally impede data collection. For example, a participant might refuse or revoke permissions, force the app to close or uninstall the app during the data collection process (Geyer, Ellis and Piwek, 2020). Usage Logger can generate crash reports (via PDFs that are exported with raw data) that include details regarding of the manufacturer/type of phone, the section of code that caused the error and a timestamp. Following a crash, the app will restart and ask participants if they are willing to continue. This feature is included to ensure that the app does not keep repeatedly crashing and instead requests that participants should contact the researchers or developers in the first instance if problems persist.

4.7 Validity

While it is not feasible to test software on every version of Android OS running on a variety of physical devices, throughout development we wanted to ensure that Usage Logger can accurately collect data from the majority of smartphones in circulation. In line with our aim to ensure that Usage Logger is straightforward to use and collects data as intended, we engaged with three separate strands of validation that transition from qualitative logging and real-world user testing to the development of highly controlled, automated systems to confirm accurate logging. The information gathered throughout supported the development and optimization of the app and additional resources.

1. *Log Books*

Throughout development and testing, researchers used pen and paper logbooks to ensure that actions performed on a given device were recorded by Usage Logger as expected. This process was repeated with each iteration of development to ensure functionality remained consistent.

2. Real-world Testing with Participants

An earlier version of Usage Logger (Activity Logger) was tested in the field to ensure usability and validity. The resulting data from these tests are reported as part of Shaw et al. (2020). Using similar techniques, the app was set up to listen to three specific events: a phone being turned on, the screen being activated, and the screen being turned off. Participants who completed this earlier study (N=46) reported no issues when installing or using the app and were asked to keep their phone switched on for the duration of the study (several days)[2]. Participants were also provided with visualisations of their usage patterns after taking part and were able to recognise repetitive patterns of daily usage. For example, when using their smartphone as an alarm clock, a regular marker of usage was repeated at the same point in time every day.

3. Software Validation

Finally, we conducted a series of automated validations with additional software. This involved running a separate app (Psych Validator – see supplementary materials for link), which programmatically causes a smartphone to perform specific actions (e.g., change app, send notification) or prompts a user to perform a particular action (turn on/off screen, un/install app). This app also documents the time at which these actions occur. For actions that were automated, Psych Validator listens in the same fashion as Usage Logger and performs an additional check to confirm an event has occurred. To assess if the app was accurately documenting the un/installation of apps, it would recount the number of apps which are

[2] As a further sense-check, Shaw et al., (2020) observed similar relationships between high-level objective usage (e.g., total smartphone time) and health assessments (e.g., depression) irrespective of whether technology use was extracted from an early version of Usage Logger (N=46) or using Apple Screen Time (N=199).

installed after the un/installation and check if the anticipated number of apps matched previous records. Usage Logger was customized so that the retrospective logging occurred after continuous logging so both the types of data can be assessed against the validation data.

4.8. Method

4.8.1 Procedure

We tested three popular Android smartphones from different manufacturers (Nokia, Huawei, & Google), which were running version 8 or later of the Android operating system. Usage Logger was installed and permissions were enabled so that continuous logging would be running the background. We then installed and started the validation app (Psych Validator).

This app automatically ran a set number of events: 20 screen ON/OFF’s, two identical app installations and two app un-installs. App events were also initiated by the validation application: 10 notifications would be pushed and removed, and a new app was opened 20 times. Data was then exported from Psych Validator and Usage Logger to the researcher’s email. Time stamps of events prompted by the validator were aligned with recorded events from Usage Logger. The differences between average time stamps were then computed.

4.9 Results

Our results confirm accurate functionality of the app to within a few seconds (Tables 2 and 3). All actions were correctly detected, but not all the attributes were captured at the exact time they occurred. Errors are reported in milliseconds.

[Table 4.2. Descriptive statistics showing discrepancies (in milliseconds) between Usage Logger (continuous logging) and Psych Validator [Usage Logger Timestamp-Psych Validator timestamp]]

Device		Nokia		Huawei		Pixel	
Event	n	M	SD	M	SD	M	SD
Screen off	10	-732.8	21.5	342.9	24.37	-523.1	93.4
Screen on	10	-476.2	9.5	342.1	15.39	-502.2	157.4
App change	20	563.6	406.9	-557.4	334.1	523.6	253.1
Notification generated	10	114.9	22.1	184.6	10.58	332.9	589.3
Notification removed	10	12.5	14.7	227.8	13.17	34.5	91.5
App installed	2	-665	1652	-636.5	121.5	-2302.5	1371.5
App uninstalled	2	2182	1117	907	19	-1578.5	1368.5

M = mean, *SD* = standard deviation

[Table 4.3. Descriptive statistics for discrepancies (in milliseconds) between Usage Logger (retrospective logging) and Psych Validator [Usage Logger Timestamp-Psych Validator timestamp]

Device		Nokia		Huawei		Pixel	
Event	n	M	SD	M	SD	M	SD
Screen off	10	1666.4	1662.8	-1026.4	28.7	-2535.1	2145.2
Screen on	10	-313.9	1066.8	-617.2	46.3	1217.9	1118.2
App change	20	316	148.9	57.4	18	-470.6	1101.3

M = mean, *SD* = standard deviation

At a millisecond time resolution, it appears that some actions were actually detected a fraction of a second before they occurred. In these instances, Usage Logger appears to be predicting the future. Of course, this is not possible, but a consequence of how Android and other operating systems run multiple programmes across physical processors (Novac, Novac, Gordan, Berczes and Bujdosó, 2019). While it appears that multiple programmes are operating in unison this is an illusion. Android maintains a list of all programs currently running and swaps between them quickly so that users perceive them to be running simultaneously. Programs swap in and out of being executed in the order of every few milliseconds, but the order in which programs are swapped in and out of being executed will vary depending on a variety of other factors including task priority, which will be determined based on other background and foreground processes (He, Chen, Wang, Wu and Yan 2019).

When an event occurs that Usage Logger records, it is possible that Android will let Usage Logger know the event has happened before it lets another app deal with the event itself. For example, when a "screen on" event occurs, the first part of Android to know that "the screen is going to be turned on" is called the Kernel. The Kernel does two things with this information: 1) it adds the "screen on" event to the list of logged events which need to be processed; 2) it adds the command "turn on the screen" to another list of things it needs to do in the near future. Having logged what has occurred the Kernel then decides what to do next. It could choose to actually turn on the screen, or to swap in the Usage Logger application (which will record the event) or do something else entirely. There are no guarantees about what happens first and so the "screen on" event could be recorded before the screen actually turns on, or vice versa. This effect also varies across devices. However, this variance operates in a fashion that will not affect the results of any investigation. If we removed some of the precision from our timing measurements, the effect would disappear completely however, we feel it is important to acknowledge these limitations here as part of a complete validation.

Overall, the quality of the data is high and suitable for the majority of research purposes that do not rely on millisecond level accuracy. We recognize that this specific form of validation represents a very small number of smartphones across a few Android operating system versions. Researchers can of course conduct their own validations as we have provided access to the source code of Psych Validator and made the app available via the Google Play Store (see supplementary materials). Alternatively, participants could, at random intervals, report what they were last using their device for via ecological momentary assessment, which could be compared with objective logs. However, this again relies on participants correctly remembering individual technology interactions, which previous research suggests is far from accurate (Andrews et al., 2015).

4.10 Data Processing and Analysis

After reliable and valid data has been collected, a second website (see supplementary materials) has been developed to help researchers decrypt participant data easily. These tools are also open source and can be developed further by other researchers. JavaScript, run from within the provided website enables the decrypting of pdfs using pdf.js (Mozzila 2019), which allows for the rendering of text within while retaining its position. This helps produce an easily interpretable .csv file. Alternatively, data can be decrypted without this website for example, by using the pdfTools package in R (Ooms, 2020) however, some tools can occasionally struggle to separate different cells in a PDF table.

This remainder of this section will walk through the process of analysing example data provided in our supplementary materials. The data will be processed, informative variables computed and a simple data visualisation generated. The included python notebook (Jupyter) and R scripts replicate these calculations and visualisations.

4.10.1 Data processing

Many data processing decisions that relate to collected data will be dependent on specific research questions. First, the researcher must choose which events are relevant. These can include: app moved to foreground, app moved to background, user interaction occurred, etc. (Android, 2019a). We have left space for researchers to dictate this in the included scripts. Recorded events can include actions which a participant had no control over (e.g., configuration changed, flush to disk, standby bucket changed). A second stage of processing involves removing any duplications (if required). Duplications are more likely to appear in retrospective datasets (not during continuous logging) where the Android operating system is responsible for developing and curating the dataset. However, if the same event was documented as occurring twice within a few milliseconds, we can be certain that the duplication is a simple double count issue. We also note that in most other instances, repetitive behaviour is common for the majority of smartphone users in the general population (Shaw et al., 2020). Following this processing, multiple types of interaction can be extracted into a single data frame. Researchers can also remove (or flag) events that appear to be the result of a participant simply leaving their phone on rather than actively using their device. For instance, a participant having a clock app in the foreground for multiple hours may be due to a participant having their device's screen set to remain on during charging (note: duration of event must be established before this processing can occur). After cleaning retrospective data specifically, it should be in a relatively similar format as continuous data and can be utilised to compute identical variables.

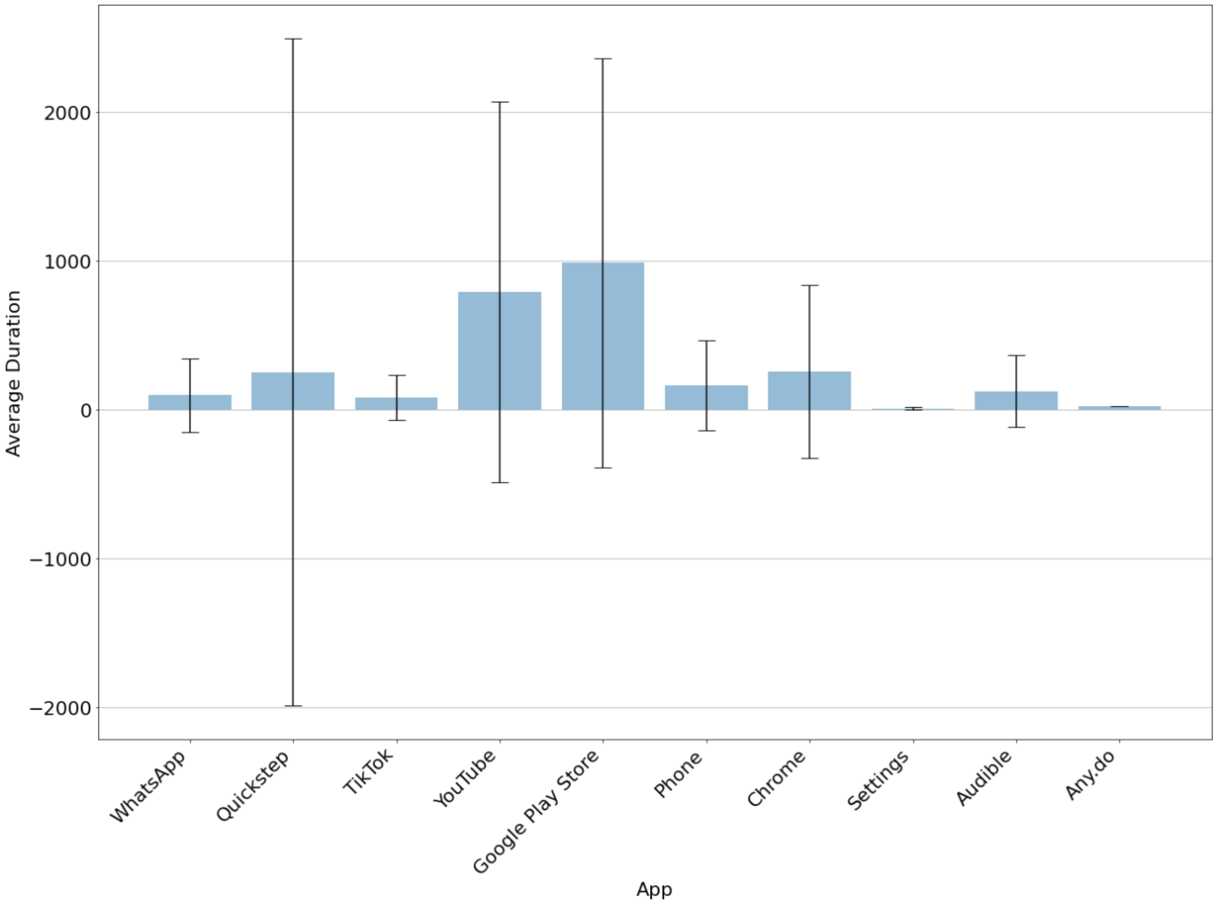
4.10.2 Establishing informative variables

The sheer number of potential variables is beyond the scope of this article. However, the supplementary code extracts a commonly referred to, but rarely measured metric, specifically, total time spent using a smartphone. Unix time stamps can be compared between a 'Screen On' event and 'Screen Off' event to calculate the duration of smartphone use. By selecting specific time frames, a variety of descriptive statistics can quantify hourly, daily, and weekly use. Similarly, establishing which apps a participant has spent more time using can be quantified by extracting app event logs, calculating the time differences between consecutive events and summing those durations independently.

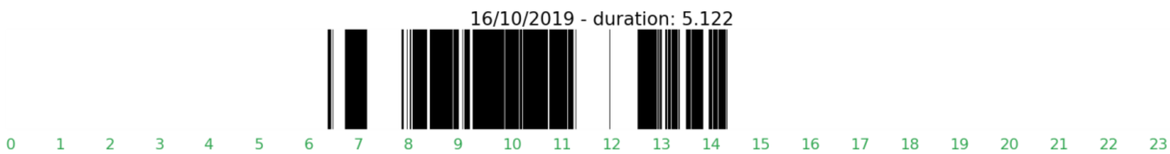
There are several other are instances where it may be advantageous to combine contextual information with usage logging. By iterating through similar subsets of contextual data, a researcher can review when an app requested specific permissions and if they were approved. This alone would provide multiple insights into understanding privacy and security from the perspective of apps or participants (Ellis 2020). Specifically, the extent to which participants approach permissions across all or specific apps installed on the phone can be explored dynamically over time.

4.10.3 Visualisation

Visualisation provides an improved descriptive understanding of usage that has been deployed as part of previous research designs (e.g., Andrews et al., 2015; Wilcockson, Ellis, & Shaw, 2018). Here we provide code to produce bar-code like visualisations. Figure 2 illustrates how much time an individual has spent using different apps over time. Figure 3 captures how a device was used across the day alongside the first author's ongoing battle with mild insomnia. Visualisations like this also help ensure an accurate representation of records and identify any errors. Figure 4 captures similar data showing the five most used apps, with the reported duration reflecting only those apps involved. These can be customized further using the provided scripts. However, the potential for other visualisations remains vast particularly in terms of identifying different patterns of use at specific points in time or understanding the flow of habitual patterns that may be goal directed or absent minded (Marty-Dugas, Ralph, Oakman & Smilek, 2018). These alone may help identify behaviors that are associated with positive, negative or neutral psychological processes and outcomes.

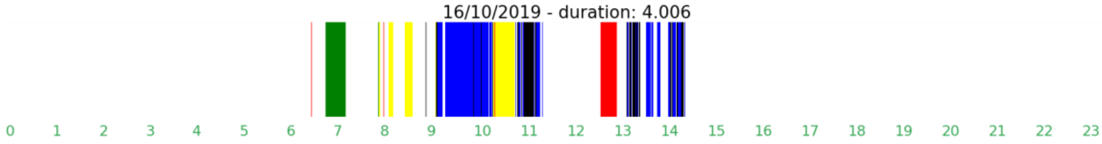


[Figure 4.2. Average durations of usage from frequently used apps in a single 24-hour period with error bars for the standard deviation.] [3]



[3] Note that the error bars are demonstrating the large variance. No negative values were recorded.

[Figure 4.3. Usage of a smartphone over a 24-hour period. Time and duration are reported. Black bars represent periods of consistent use.]



[Figure 4.4. Usage of specific apps over a 24-hour period. Colors represent different apps including: WhatsApp (Black), Quickstep (Red), TikTok (Green), YouTube (Yellow), and Google Play Store (Blue).]

4.11 Discussion and Conclusion

When it comes to understanding the impact of mass communications technology on individuals and groups, psychology's current conclusions are only as strong as the measurements that underpin any design. The same notion applies to large swathes of research that aims to understand the predictive properties of technology use across psychological science (Ellis et al., 2019; Ellis, 2020). Existing measures have been more informed by concern around technology use (e.g., smartphone 'addiction') rather than making the most of technological resources at the disposal of behavioral scientists (Ellis, 2020). Several indications suggest that the relationship between smartphone usage and well-being has been overestimated when relying on objective data (Ellis et al., 2019; Katevas, Arapakis & Pielot, 2018; Shaw et al., 2020).

The method documented here securely and accurately provides detailed accounts of smartphone usage We acknowledge that further work could ensure that the security of participant data is enhanced further. For example, emailing a password to a researcher is only as secure as a researcher's email account. However, it is worth noting that while high-level raw data (e.g., total smartphone time) in this instance is unlikely to compromise an individual's safety or security if it were widely available, data regarding specific apps could be used to make inferences about individuals that they may wish to keep private (Ellis, 2020). Researchers should be especially mindful when linking digital traces like these with other psychological assessments or sensitive demographic variables. As a consequence, additional security procedures might include uploading data to a secure server or existing cloud service, but this might increase the technical threshold for adoption. Our decision to make Usage Logger serverless was also to ensure that researchers can use the app and comply with open science practices from the outset, but we acknowledge these benefits can generate conflicts with privacy requirements (Dennis et al., 2019). If Usage Logger was developed purely from a software security perspective then its architecture would be very different however in this instance, security and privacy decisions meet our original aims and ensure participants remain in control of their data (Dennis et al., 2019; Geyer, Ellis & Piwek, 2020). This is similar to how Apple allow iPhone users to export all their Health app data in the form of Extensible Markup Language (XML) file. However, unlike Usage Logger, the data is extensive, sensitive, and not encrypted.

In many respects, Usage Logger is really only the start of a code base that could be diversified further in order to collect data across multiple devices and services that capture technology-related behaviors. In addition, the digital data generated from related methods has many more applications beyond those already discussed. This includes researchers going beyond device or application-level metrics. For example, Meier and Reinecke (2020) consider different types of interaction, the device level (e.g., time spent on a smartphone), the application level (e.g., time using a specific app), or the feature level (e.g., using specific features within Twitter). Our tools allow for a complete analysis at a device and application level. Some feature level analysis is possible based on specific app notifications. However, while measuring technology behaviors on a more granular level will provide additional insights, technology companies will need to have transparent, accessible APIs and access points for researchers to investigate in-app behavior (Johannes, Vuorre & Przybylski, 2020). However, the tools reported here do provide access to metrics that will help researchers working across a variety of domains. This includes the ability to describe smartphones interactions and better understand their impacts. Enabling descriptive work at any level remains essential in order to aid with the development of well-grounded theory, which remains a long-standing aspiration for those studying the causal effects of new technology on people and society (Ellis, 2019; Ellis, 2020; Miller, 2012).

These developments are not to suggest that there is no place for non-behavioral measures in this domain of research. On the contrary, if a research question aims to consider a persons' thoughts, feelings, or attitudes towards a specific technology then other measures will remain essential. However, psychometric tools to support this endeavour should be developed and used with a clear appreciation of the specific questions they can (and cannot) answer. For example, many survey instruments are not an accurate reflection of objective usage despite often being used as a proxy for behaviors (Ellis et al., 2019). Assuming that technology use is the primary variable of interest, researchers may consider moving away from latent measures completely given that 'use' is directly observable (Ellis, 2020).

When it comes to behavior, no self-report measurement will be perfect compared with ground truth (Orben, Dienlin & Przybylski, 2019). However, this ground truth is slowly becoming readily available and we would encourage scientists to adopt these methods alongside open research practices wherever possible. While not standard practice for those who often make sizable claims about the effects of technology on large swathes of the population, combing such an approach with novel analytical methods are essential for the field to progress.

Only then can an interdisciplinary endeavour deliver valuable insights for both scientists and policymakers (Ellis, 2020).

Open Practices Statement

All source code, materials and data are available (see supplementary materials)

References

Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. (2011). Social fMRI: investigating and shaping social mechanisms in the real world. *Pervas Mobile Comput.* 7, 643–659

Aledavood, T., Lehmann, S., & Saramäki, J. (2018). Social network differences of chronotypes identified from mobile phone data. *EPJ Data Science*, *7*(1), 18–20.

Alexander-Bown, S. (2019, April 27). Secure preferences. Retrieved April 29, 2019, from https://github.com/scottyab/secure-preferences

Alimomeni, M. (2014). *New Notions of Secrecy and User Generated Randomness in Cryptography* (Doctoral dissertation, University of Calgary).

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: tools to compare estimated and real-world smartphone use. *PLOS ONE*, *10*(10), e0139004.

Android. (2019a). UsageEvents.event [computer software]. Retrieved April 29, 2019, from https://developer.android.com/reference/android/app/usage/UsageEvents.Event

Android. (2019b). Save files on device storage [computer software]. Retrieved April 29, 2019, from https://developer.android.com/training/data-storage/files

Android. (2019c). FileProvider [computer software]. Retrieved April 29, 2019, from https://developer.android.com/reference/android/support/v4/content/FileProvider

Apple (2019) Screen Time (version unknown) [Mobile application software]

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396–403.

Barthelmäs, M., Killinger, M., & Keller, J. (2020). Using a Telegram chatbot as cost-effective software infrastructure for ambulatory assessment studies with iOS and Android devices. *Behavior Research Methods*, 1-8.

Bellovin, S. M. & Merritt, M. (1992). Encrypted key exchange: password-based protocols secure against dictionary attacks. In *Proceedings 1992 IEEE Computer Society Symposium on Research in Security and Privacy* (pp. 72–84). IEEE.

Boase, J., & Ling, R. (2013). Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*, *18*(4), 508–519.

Clayton, R. B., Leshner, G., & Almond, A. (2015). The extended iSelf: The impact of iPhone separation on cognition, emotion, and physiology. *Journal of Computer-Mediated Communication, 20*(2), 119-135.

Davidson, B. I., Ellis, D., Bowman, N., Liveley, G., Shaw, H., Przybylski, A. K., & Levine, M. (2019). Avoiding Irrelevance: The manifestation and impacts of technophobia in psychological science.. *PsyArXiv* https://doi.org/10.31234/osf.io/b9f4p

Dennis, S., Garrett, P., Yim, H., Hamm, J., Osth, A. F., Sreekumar, V., & Stone, B. (2019). Privacy versus open science. *Behavior Research Methods, 51*(4), 1839-1848.

Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin, 13*, e25025.

Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders, 207*, 251-259.

Ellis, D. A. (2019). Are smartphones really that bad? improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior, 97*, 60-66.

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior?. *International Journal of Human-Computer Studies, 130*, 86-92.

Ellis, D. A. (2020). *Smartphones within Psychological Science*. Cambridge University Press

Facebook (2019) Messenger (236.0.0.14.120) [Mobile application software]

Ferreira, D., Kostakos, V., and Dey, A. K. (2015). AWARE: mobile context instrumentation framework. *Frontiers in ICT*, *2*(6), 1-9.

He, J., Chen, T., Wang, P., Wu, Z., & Yan, J. (2019, December). Android multitasking mechanism: Formal semantics and static analysis of apps. In *Asian Symposium on Programming Languages and Systems* (pp. 291-312). Springer, Cham.

Hinds, J., & Joinson, A. (2019). Human and computer personality prediction from digital footprints. *Current Directions in Psychological Science*, *28*(2), 204-211.

Johannes, N., Meier, A., Reinecke, L., Ehlert, S., Setiawan, D., Walasek, N., Dienlin, T., Buijzen, M. & Harm, V. (2019). The Relationship Between Online Vigilance and Affective Well-Being in Everyday Life: Combining Smartphone Logging with Experience Sampling. *PsyArXiv*, https://doi.org/10.31234/osf.io/b9f4p

Johannes, N., Vuorre, M., & Przybylski, A. K. (2020). Video game play is positively correlated with well-being. *PsyArXiv*, https://doi.org/10.31234/osf.io/qrjza

Katevas, K., Arapakis, I., & Pielot, M. (2018, September). Typical phone use habits: Intense use does not predict negative well-being. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (p. 11). ACM.

Keil, T. F., Koschate, M., & Levine, M. (2020). Contact Logger: Measuring everyday intergroup contact experiences in near-time. *Behavior Research Methods*, 1-19.

Lepp, A., Barkley, J. E., & Karpinski, A. C. (2015). The relationship between cell phone use and academic performance in a sample of US college students. *Sage Open*, *5*(1), 2158244015573169.

Marty-Dugas, J., Ralph, B. C., Oakman, J. M., & Smilek, D. (2018). The relation between smartphone use and everyday inattention. *Psychology of Consciousness: Theory, Research, and Practice, 5(*1), 46.

Meier, A. and Reinecke, L. (2020). Computer-mediated communication, social media, and mental health: A conceptual and empirical meta-review. *Communication Research*, 1–28.

Nelson, D. & Vu, K.-P. L. (2010). Effectiveness of image-based mnemonic techniques for enhancing the memorability and security of user-generated passwords. *Computers in Human Behavior*, *26*(4), 705–715.

Novac, O. C., Novac, M., Gordan, C., Berczes, T., & Bujdosó, G. (2017, June). Comparative study of Google Android, Apple iOS and Microsoft Windows phone mobile operating systems. In 2017 *14th International Conference on Engineering of Modern Electric Systems (EMES)* (pp. 154-159). IEEE.

Ooms, J. (2020). Pdftools [Computer software]. Retrieved from: https://cran.r-project.org/web/packages/pdftools/pdftools.pdf

Orben, A., & Przybylski, A. K. (2019a). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173.

Orben, A., & Przybylski, A. K. (2019b). Screens, teens, and psychological well-being: evidence from three time-use-diary studies. *Psychological Science*, *30*(5), 682-696.

Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Reply to Foster and Jackson: Open scientific practices are the way forward for social media effects research. *Proceedings of the National Academy of Sciences*, *116*(31), 15334-15335.

Piwek, L., Ellis, D. A., & Andrews, S. (2016). Can programming frameworks bring smartphones into the mainstream of psychological science?. *Frontiers in Psychology*, *7*, 1252.

Pliam, J. O. (2000, December). On the incomparability of entropy and marginal guesswork in brute-force attacks. In *International Conference on Cryptology in India* (pp. 67-79). Springer, Berlin, Heidelberg.

Richardson, M., Hussain, Z., & Griffiths, M. D. (2018). Problematic smartphone use, nature connectedness, and anxiety. *Journal of Behavioral Addictions*, *7*(1), 109-116.

Reinfelder, L., Schankin, A., Russ, S., & Benenson, Z. (2018, September). An Inquiry into Perception and Usage of Smartphone Permission Models. In *International Conference on Trust and Privacy in Digital Business* (pp. 9-22). Springer, Cham.

Rosen, L. D., Lim, A. F., Felt, J., Carrier, L. M., Cheever, N. A., Lara-Ruiz, J. M., ... & Rokkum, J. (2014). Media and technology use predicts ill-being among children, preteens and teenagers independent of the negative health impacts of exercise and eating habits. *Computers in Human Behavior*, *35*, 364-375.

Rosen, L., Carrier, L. M., Miller, A., Rokkum, J., & Ruiz, A. (2016). Sleeping with technology: cognitive, affective, and technology usage predictors of sleep problems among college students. *Sleep Health*, *2*(1), 49-56.

Shaw, H., Ellis, D. A., Geyer, K., Davidson, B. I., Ziegler, F. V., & Smith, A. (2020). Quantifying smartphone 'use': Choice of measurement impacts relationships between 'usage' and health, *Technology, Mind, and Behavior*, *1* (2).

Sassenberg, K., & Ditrich, L. (2019). Research in Social Psychology Changed Between 2011 and 2016: Larger Sample Sizes, More Self-Report Measures, and More Online Studies. *Advances in Methods and Practices in Psychological Science*, 2515245919838781.

SnapInc. (2019). Snapchat - android apps on google play. Retrieved April 29, 2019, from https://play.google.com/store/apps/details?id=com.snapchat.android

Turner, A., Topor, M., Stewart, A. J., Owen, N., Kenny, A. R., Jones, A., & Ellis, D. (2020). Open Code/Software: A primer from UKRN.

Van Vliet, H., (2008). *Software engineering: principles and practice*. Hoboken, NJ: John Wiley & Sons.

Wilcockson, T. D., Ellis, D. A., & Shaw, H. (2018). Determining typical smartphone usage: what data do we need? *Cyberpsychology, Behavior, and Social Networking*, *21*(6), 395–398.

Xu, R., Frey, R. M., Fleisch, E., & Ilic, A. (2016). Understanding the impact of personality traits on mobile app adoption–insights from a large-scale field study. *Computers in Human Behavior, 62*, 244–256.

Zetetic. (2019). SQLCipher for android - zetetic. Retrieved April 29, 2019, from https://www.zetetic.net/sqlcipher/sqlcipher-for-android/

Supplementary Materials

Links

App - https://play.google.com/store/apps/details?id=geyerk.sensorlab.suselogger

App source code and associated websites - https://github.com/kris-geyer/UsageLoggerPublished

Psych Validator - https://github.com/kris-geyer/psychvalidaitor

Walkthrough Guide - https://u-log-walk.netlify.app/

Funding

This work was part funded by the Centre for Research and Evidence on Security Threats [ESRC Award: ES/N009614/1]

Acknowledgements

We thank Ben Marshall for comments on an earlier version of this manuscript

Chapter IV

Do smartphone usage scales predict behaviour?

Chapter Publication status: Published

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior?. *International Journal of Human-Computer Studies, 130*, 86-92.

Specific aim

2. Establish if there is a theoretical grounding for the use of psych apps relative to conventional methodologies.

My contribution

I contributed to discussions of the research and the general research design, conducted the statistical analysis in parallel with other researchers to ensure the accuracy of the findings, edited drafts of the work and added additional analysis such as controlling for multiple comparisons.

The previous two chapters documented how multiple open-source Android apps can operate in a psychological context. The apps demonstrated that collecting different streams of data require different considerations for researchers and participants. Additionally, the inclusion of external sensors is complex but manageable.

While these methodological contributions deliver insights about the development of new methods alongside logistical considerations, questions about their potential utility for psychological science remain. Can such apps provide meaningful contributions? Does the use of such apps highlight limitations in long standing practices of social psychology?

To explore this further, this Chapter compared 238 iPhone users' estimates of smartphone usage and objective records of their own smartphone usage behaviour. We also make comparisons with validated scales of smartphone addiction to review if there is any relationship between amount of smartphone usage and the degree that scales reported participants being addicted. This chapter begins to examine the question: are questionnaires about behaviour sufficiently accurate to be comparable with objective records of behaviour? Currently, we only use Apple screen Time to capture data on smartphone usage but *Chapter VI* then utilises an app which was the product of *Chapter IV* to far more transparently capture such data. As we were developing the psych apps, Apple released the screen time app. For this reason we used this system during this study. I encourage any researchers to take the publicly available usage logger and attempt to replicate our findings with android data.

Questionnaires on smartphone addiction were found to have a very small correlation with smartphone usage. After the publication of this paper in 2019, the work was reported in New Scientist magazine (Chivers, 2018), and the New Statesman (Chivers, 2019). Findings were even mentioned as a part of the government investigation into smartphone usage (UK Parliament, 2019). This study is currently the 2nd most cited paper in the journal in the previous three years. It currently has been cited 146 times. The post-print has been downloaded nearly 2,500 times.

Abstract

Understanding how people use technology remains important, particularly when measuring the impact this might have on individuals and society. However, despite a growing body of resources that can quantify smartphone use, research within psychology and social science overwhelmingly relies on self-reported assessments. These have yet to convincingly demonstrate an ability to predict objective behavior. Here, and for the first time, we compare a variety of smartphone use and ‘addiction’ scales with objective behaviors derived from Apple’s Screen Time application. While correlations between psychometric scales and objective behavior are generally poor, single estimates and measures that attempt to frame technology use as habitual rather than ‘addictive’ correlate more favorably with subsequent behavior. We conclude that existing self-report instruments are unlikely to be sensitive enough to accurately predict basic technology use related behaviors. As a result, conclusions regarding the psychological impact of technology are unreliable when relying solely on these measures to quantify typical usage.

Introduction

1.1Background

Despite decades of progress, understanding the overall impact of technology on people and society remains a challenge (Shaw et al., 2018). Perhaps this is because such a topic naturally aligns itself with many disparate research questions. Investigations range from issues concerning problematic use (e.g., can smartphones disrupt sleep?), to the effects of engaging with feedback as part of a behavior change intervention (e.g., does monitoring physical activity improve health?) (Ellis & Piwek, 2018). Approaches to date in behavioral science have almost exclusively focused on asking people to consider their personal experience with a technology in order to better understand its impact (Ellis, Kaye, Wilcockson, & Ryding, 2018). This mirrors a general trend within social psychology as a whole (Baumeister, Vohs, & Funder, 2007; Dolinski, 2018), but it is perhaps more surprising when applied to mobile and pervasive systems that can record human-computer interactions directly (Piwek, Ellis, & Andrews, 2016). Smartphones have provided several new opportunities in this regard (Miller, 2012). For example, behavioral interactions can be measured 'in situ 'with a variety of applications and those in computer science have been measuring these interactions for several years (Jones, Ferreira, Hosio, Goncalves, & Kostakos, 2015; Oliver, 2010; Zhao et al., 2016). However, methodological developments have had very little impact on how the majority of social science attempts to quantify, explain, and understand technology use more generally.

Two common methods are often deployed by social scientists to capture technology usage 'behaviors'. The first relies on participants providing estimates of frequency or duration (Butt & Phillips, 2008). However, this method has previously been described as 'sub-optimal ' when attempts are made to validate single measures against objective behavior (e.g., Boase & Ling, 2013). In addition, the use of multiple technologies simultaneously (e.g., a smartphone and a laptop) mean that these estimates have become even more problematic due the level of cognitive burden required to quantify many different types of habitual behavior (Boase & Ling, 2013; Doughty, Rowland, & Lawson, 2012; Jungselius & Weilenmann, 2018). In response to these critiques, a second method utilizes questionnaires that aim to quantify technology related experiences. Considering smartphones specifically, an abundance of self-reported measures have been created in an attempt to capture and predict actual behavior (e.g., Bianchi & Phillips, 2005; Billieux, Van Der Linden, & Rochat, 2008; Csibi, Demetrovics, & Szabó, 2016; Kwon, Kim, Cho, & Yang, 2013; Rosen, Whaling, Carrier, Cheever, & Rokkum, 2013; Sivadas & Venkatesh, 1995; Yildirim & Correia, 2015). Following traditional methods associated with scale development, factor analyses ensure that such assessments are reliable, but less emphasis has been placed on establishing validity. This sets these scales apart from other areas where self-report has been rigorously validated against behavioral metrics (e.g., personality) (e.g., McCrae & Costa, 1987; Parker & Stumpf, 1998). The lack of validation and clarity regarding constructs and measurement is therefore detrimental to the sound utilization of these scales in subsequent research (Clark & Watson, 1995).

Many measures are conceptualized around 'smartphone behaviors', and are used by many researchers to provide a proxy measure of usage (Ellis et al., 2018). Perhaps more importantly, research utilizing these assessments tends to use high-scores to correlate smartphone usage with a variety of negative outcomes (e.g., depression and anxiety) (e.g., Elhai, Dvorak, Levine, & Hall, 2017; Richardson, Hussain, & Griffiths, 2018) and provide evidence for the classification of a behavioral addiction (e.g., Tao et al., 2017; Wolniewicz, Tiamiyu, Weeks, & Elhai, 2018). This repeats a pattern of research priorities that previously focused on the negative impacts of many other screen-based technologies, systematically moving from television and video games, to the internet and social media (Przybylski & Weinstein, 2017; Rosen et al., 2014). However, the few studies that have measured behavior directly, tend to demonstrate conflicting results. For example, Rozgonjuk et al. (2018) observed no association between smartphone use and severity of depression or anxiety. Further, higher levels of reported depression correlated with individual's checking their phone *less* over a week. Therefore, the notion of reducing 'screen time' and technology may be counter-intuitive, as a sudden reduction in smartphone use may in fact be an early warning sign of social withdrawal (Mou, 2016).

1.2 The Present Study

To date, only a handful of small studies have attempted to validate these scales in small samples that focus on single measures with mixed results (Andrews, Ellis, Shaw, & Piwek, 2015; Elhai et al., 2018; Foerster, Roser, Schoeni, & Röösli, 2015; Lin, Chiang, & Jiang, 2015; Rozgonjuk et al., 2018; Wilcockson, Ellis, & Shaw, 2018). Here, we attempt to compare the human accuracy of ten smartphone usage scales and single estimates against objective measures of smartphone behavior. This takes advantage of a recent iOS update from Apple, which automatically logs a series of behavioral metrics related to 'screen time' over a period of seven days. Data available includes the length of time users spend on their devices, the number of times the phone is picked up, alongside the number of notifications received daily. This allowed for several attempts at validation that includes correlations and cluster-based analyses. The latter of which compares the overlap between high-usage groups derived independently from self-report scores or behavioral metrics.

Method

2.1Ethics

This study was ethically approved by the University of Bath School of Management (ID: 2392) and was conducted in accordance with guidelines provided by the British Psychological Association (BPS).

2.2 Participants

Participants were recruited from within affiliated universities (Lancaster, Bath, and Lincoln) (23.12%), or using the Prolific Academic platform (76.89%). Participants were paid a small sum for their participation via Prolific Academic (£5.34/hr) and provided informed consent. 238 participants (124 female, mean age = 31.88; *SD* = 11.19) who owned an iPhone 5 or above and had been running the latest version of iOS for at least one week were eligible to participate. Our sample size is comparatively larger than other studies that have previously attempted to validate these scales and includes data from a comparable time frame (Andrews et al., 2015; Elhai et al., 2018; Lin et al., 2015; Rozgonjuk et al., 2018; Wilcockson et al., 2018). In addition, our sample is similar to studies that utilize these scales when making links between smartphone use and other correlates, for example, Wolniewicz et al (2018), N=296 and Elhai, Levine, Dvorak, and Hall (2016), N = 308.

2.3 Procedure and Materials

All participants were directed to a Qualtrics survey hosted by the University of Lincoln. Participants first provided an estimate of how many hours and minutes they spend on their iPhone daily. They were also asked to estimate the number of notifications received daily, and how many times they pick up their device each day. The specific wording was as follows: “Please estimate how many hours and minutes you spend on your phone each day” , “Please estimate how many notifications you receive on your phone each day” and “Please estimate how many times a day you pick up and use your phone”.

Next, they completed ten scales that aim to asses smartphone usage and/or associated constructs (Table 1). Scales were selected based on their popularity (generally operationalized as their citation count) and broad range of conceptualizations (e.g., attachment, fears, ‘addictions’, etc.) and were presented at random within the survey. There are practical issues associated with having participants complete an extensive amounts of questionnaires such as dropout rates. We aimed to include enough questionnaires to capture the most influential scales and get a representative view of the field as a whole, but not so many as to frustrate the participants. The degree to which a scale was selected related to the number of citations that the scale but also the amount of effort required to complete the scale. Thus our analysis could provide insights to the degree that the area of study suffers methodological limitations. Finally, participants transferred their latest Screen Time capture data from Apple's Screen Time app to provide the actual number of hours and minutes spent on their phone, number of notifications received, and number of times they had picked up their device each day for a period of one week. Daily averages were calculated for all three behavioral metrics.

Note the data is publicly available here (https://osf.io/3w74t/) to generate the complete scores of all of the scales simply sum the associated columns. The only exception is reverse code the first item in attachment scale (columns labelled "Attachment_1"), change all values in PMPUQ that are 5 to 0 and reverse code item 4 in PMPUQ.

Mobile Phone Problem Use Scale (MPPUS)

(Bianchi & Phillips, 2005) The MPPUS is a 27-item scale designed to assess problematic usage of mobile phones, with each item scored via a Likert scale ranging from 'Not true at all '(1) to 'Extremely true '(10). Higher scores denote increased levels of problematic usage.

Nomophobia Questionnaire (NMP-Q)

(Yildirim & Correia, 2015) The NMP-Q is a 20-item designed to assess nomophobia. This is defined as a phobia of being separated from one's smartphone. Each statement is scored using a 7-point Likert scale from 'Strongly disagree '(1) to 'Strongly agree '(7). Higher scores correspond to higher nomophobia severity, where scores of <20 denote an absence of nomophobia, >20 – <60 denotes mild nomophobia, >=60 – <100 denotes moderate nomophobia, with scores >= 100 suggesting severe nomophobia.

Possession Incorporation in the Extended Self

(Sivadas & Venkatesh, 1995) This scale comprises of 6-items that aims to determine the extent possessions have become incorporate into an ‘extended self ’originally defined by Belk (1988). Statements are scored using a 7-point Likert scale ranging from ‘Strongly disagree ’(1) to ‘Strongly agree ’(7). We used the specific-possession incorporation version, where the items were phrased as follows: ‘x helps me achieve the identity I want to have’, with x substituted as ‘my smartphone,’. Higher scores denote an increased integration of a smartphone an identity.

Attachment Scale

(Sivadas & Venkatesh, 1995) The attachment scale contains 4-items, which aims to assess the attachment to an object, in this case a smartphone, for example, ‘I am emotionally attached to my smartphone’. This used a 7-point Likert scale ranging from ‘Strongly disagree ’ (1) to ‘Strongly agree ’(7). Higher scores correspond to higher levels of attachment to the object in question.

Smartphone Addiction Scale (SAS)

(Kwon et al., 2013) The SAS is a 33-item scale designed to measure smartphone ‘addiction’, with each statement scored via a 6-point Likert scale from ‘Strongly disagree ’(1) to ‘Strongly agree ’(6). It consists of six factors: daily life disturbance, positive anticipation, withdrawal, cyberspace-orientated relationship, overuse, and tolerance. These can be combined to provide a single score. Higher scores correspond to higher smartphone usage and ‘addiction’.

Smartphone Application-Based Addiction Scale (SABAS)

(Csibi et al., 2016) We used the English version of the SABAS scale, which comprises of 6-items, with each item scored using 6-point Likert scale from 'Strongly disagree '(1) to 'Strongly agree '(6). It aims to assess application-based addictions associated with smartphones. Higher scores correspond to higher smartphone (application) usage and 'addiction'.

Problematic Mobile Phone Use Questionnaire (PMPUQ)

(Billieux et al., 2008) The PMPUQ aims to assess actual and potential problematic usage of mobile phones. We used a short 15-item version, which concerned mobile phone usage when driving, forbidden use of mobile phones, and use of mobile phones in dangerous situations. The scale is traditionally a 4-item Likert scale from 'Strongly disagree '(1) to 'Strongly agree '(4), however, we also included an additional 'Not Applicable '(5) for those who did not drive in our sample (coded as 0). Higher scores correspond with increased levels of problematic usage.

Media and Technology Usage and Attitudes Scale (MTUAS)

(Rosen et al., 2013) The complete MTUAS comprises of 66-items that aims to assess technology and media use more widely. However, here we used 9-items from a subscale, which focuses on smartphone use (items 9-17). Each item is scored on a 10-point scale from 'Never ' (1) to 'All the time '(10), where the mean measure is taken for each participant. Higher means correspond to higher smartphone usage.

Smartphone Use Questionnaires (SUQ-G&A)

(Marty-Dugas, Ralph, Oakman & Smilek, 2018) SUQ-G&A seeks to distinguish general smartphone usage and absent-minded smartphone usage. This provides scores from two 10-item scales: general (SUQ-G) and absent-minded (SUQ-A). Both use a 7-point scale from 'Never '(1) to 'All the time '(7). SUQ-G focusses on specific uses, e.g., 'How often do you check social media apps such as Snapchat, Facebook, or Twitter', and the SUQ-A asks questions regarding mindless usage, e.g., 'How often do you find yourself checking your phone without realizing why you did it?'. Higher mean scores correspond to higher smartphone usages (general or absent-minded).

2.4 Analysis Plan

Scores for each scale were calculated (as detailed above), with manipulations for reversed items as necessary. Tables 1 and 2 provide descriptive statistics for all self- reported and behavioral metrics. Pearson's Correlations (Table 3) were calculated between all self-reported measures, single estimates, and objective behavioral metrics. The tests have been adjusted for multiple comparisons using the Benjamini-Hochberg (Benjamini, & Hockberg, 1995) procedure. The associated false discovery rate has been set at .05 and therefore we would expect only 5% of the statistical tests carried out are incorrectly reported as significant. While we note that the average number of notifications is not strictly a behavioral measure, it is included here to provide context regarding how often a person may be expected to pick up or check their phone as notifications act as a request for user attention. Therefore, this provides an additional validity check as we expect to observe a positive correlation between the number of notifications and the amount of time a person spends on their phone. The overall performance of each self- report measure was derived from the mean correlation across all three objective behavioral measures (Figure 1). For example, the mean score for a single duration estimate was based on mean of three correlations between the estimate and behavioral averages of (1) hours use, (2) pickups, and (3) notifications. Finally, a series of k- means algorithms considered overlaps in classification when participants were clustered using only self-report or objective behavior (Figure 2).

3 Results

3.1 Self-Reported Measures

Table 1 reports the means, standard deviations, and internal consistency measures (Cronbach's Alpha (α) for all self-reported measures.

Self-report measures	**Items**	**Min-max**	***M***	***SD***	***α***
Single time estimate (minutes) (TEst)	1	–	226.6	128.37	
Single pick up estimate (PEst)	1	–	45.69	42.16	
Single notification estimate (NEst)	1	–	39.09	42.46	
Mobile phone problem use scale (MPPUS)	27	27–270	111.9	43.12	0.94
Nomophobia scale (NS)	11	20–140	82.57	25.76	0.96
Possession incorporation in the extended self (ES)	6	6–42	21.53	8.99	0.93
Smartphone attachment scale (SAt)	4	4–24	17.02	6.05	0.87
Smartphone addiction scale (SAS)	33	33–198	94.2	30.17	0.95
Smartphone application-based addiction scale (SABAS)	6	6–36	15.83	5.89	0.81
Problematic mobile phone use questionnaire (PMPUQ)	15	15–60	27.54	5.85	0.72

Media and technology usage and attitudes scale (MTUAS)	9	9–90	6.24	1.33	0.84
Smartphone use questionnaire (general) (SUQ-G)	10	10–70	48.45	8.89	0.78
Smartphone use questionnaire (absent minded) (SUQ-A)	10	10–70	45.6	14.37	0.95

Table 4.1. Descriptive Statistics (means (M) and standard deviations (SD)) for single estimates and self-report assessments. Highest and lowest possible scores for each measure are provided for reference.

3.2 Behavioral Metrics

Table 2 presents means and standard deviations from objective behavioral measures. Data were available for the previous seven days, however, the day of data collection is naturally incomplete, so all behavioral metrics are based on an average from six complete days of data from each participant. Previous research has suggested that identical smartphone usage collected for a minimum of five days will reflect typical weekly usage, with habitual checking behaviors (pickups) requiring a minimum of

two complete days of collection irrespective of weekday (Wilcockson et al., 2018). A series of one-way ANOVAs confirm that no weekday differences were present in any of our behavioral data (all *p*'s > .2). Finally, we note that participants, on average, pickup their phones fewer times when compared to the number of notifications received (1:1.05 ratio of pick ups to notifications).

Behavioral Measure	***M***	***SD***
Time (minutes)	232.66	119.44
Pickups	85.84	53.34
Notifications	90.13	88.86

Table 4.2. Descriptive Statistics for Behavioral Measures (means (M) and standard deviations (SD)). These are in line with previous research considering smartphone behaviors in smaller samples (e.g., Andrews et al., 2015).

3.3 Correlations

Pearson's correlation coefficients were calculated across single estimates, self- reported scales, and behavioral data (Table 3). All self-reported scales positively correlated with objective time spent on a smartphone (ObjT). These varied from .40 to .13. However, a single estimate of time (TEst) was a better predictor than any self- report scale [*r* =.48].

Average number of objective pickups (ObjP) modestly correlated with the Smartphone Usage Questionnaire - General (SUQ-G) [*r* = .31] and Smartphone Usage Questionnaire – Absent Minded (SUQ-A) [*r* = .30]. Weak correlations were observed between the Smartphone Addiction Scale (SAS) [*r* = .22], Mobile Phone Problem Use Scale (MPPUS) [*r* = .18], and Media and Technology Usage and Attitudes Scale (MTUAS) [*r* = .15]. Again, a single estimate of pickups (PEst) was a superior predictor in comparison to any self-report instrument [*r* = .32].

Average number of notifications (ObjN) weakly correlated with most self-reported scales (exceptions are the Extended Self (ES), Smartphone Application Application- Based Addiction Scale (SABAS), and the Problematic Mobile Phone Use Questionnaire (PMPUQ)). These varied from .28 to .15. A single estimate of daily notifications received (NEst) correlated moderately with the objective counterpart (ObjN) [r = .53].

Table 4.3. Pearson's correlations between single estimates, self-reported scales, and objective behavior.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Age															
2. TEst	−0.22**														
3. PEst	−0.10	.22**													
4. NEst	−0.15*	.30**	.32**												
5. MPPUS	−0.08	.28**	.14*	0.06											
6. NS	−0.03	.22**	0.08	0.06	.74**										
7. ES	.14*	.14*	0.07	0	.53**	.56**									
8. SAt	0.02	.21**	0.04	0.03	.46**	.54**	.69**								
9. SAS	−0.08	.29**	0.09	0.06	.82**	.75**	.62**	.59**							
10. SABAS	−0.03	.21**	0.13	0.05	.77**	.68**	.55**	.52**	.76**						
11. PMPUQ	−0.04	.27**	.17**	.14*	.55**	.46**	.38**	.37**	.56**	.48**					
12. MTUAS	−0.26**	.28**	.24**	.22**	.36**	.38**	.23**	.32**	.34**	.25**	.37**				

13. SUQ-G	−0.28**	.36**	.14*	.24**	.56**	.54**	.39**	.41**	.57**	.43**	.42**	.60**			
14. SUQ-A	−0.26**	.24**	.14*	0.04	.66**	.58**	.35**	.40**	.62**	.53**	.47**	.45**	.69**		
15.ObjT	−0.20**	.48**	0.1	.13*	.33**	.32**	.21**	.32**	.40**	.26**	.27**	.26**	.34**	.36**	
16. ObjP	−0.32**	.23**	.23**	.32**	.18**	.16*	−0.01	0.1	.22**	0.12	.15*	.24**	.31**	.30**	.39**
17. ObjN	−0.35**	.27**	.13*	.53**	.14*	.19**	0.05	.15*	.18**	0.08	0.12	.22**	.28**	.21**	.37**

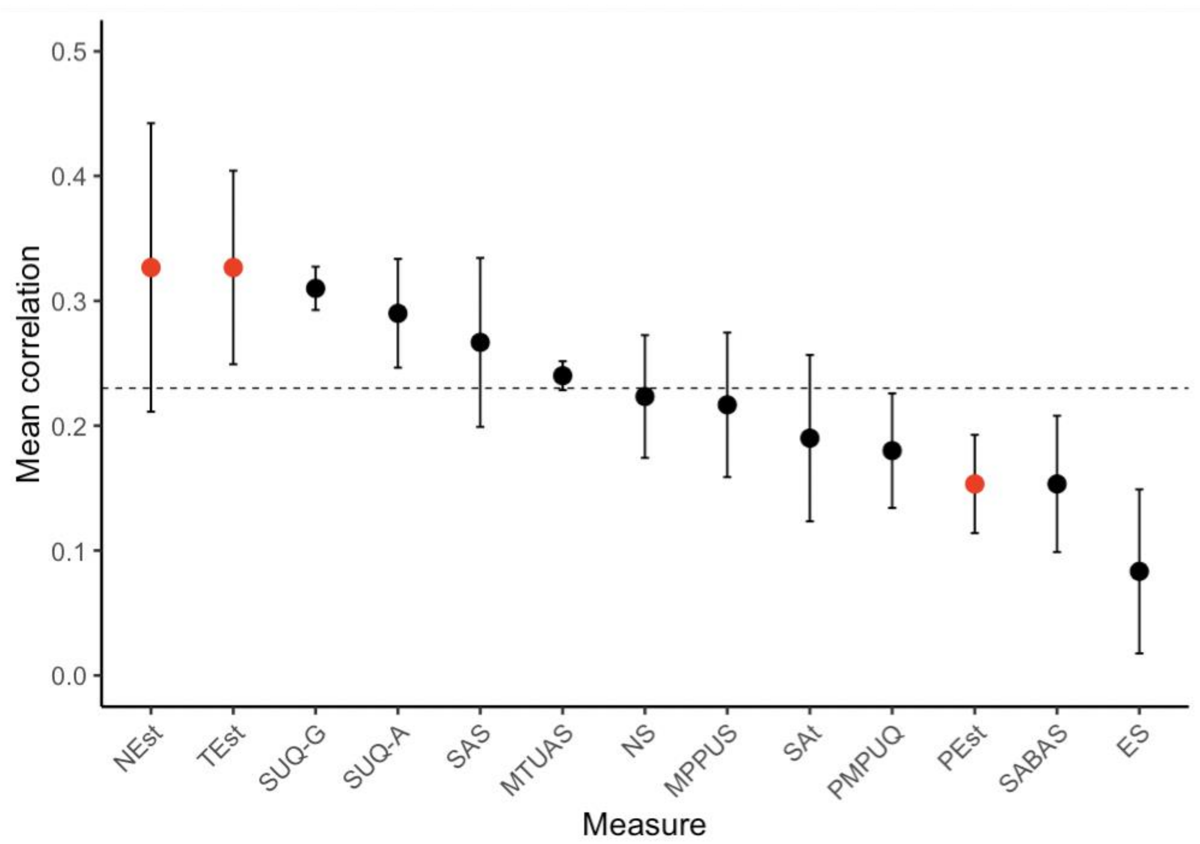
TEst = Single time estimate, PEst = Single pick up estimate, NEst = Single notification estimate, MPPUS = Mobile phone problematic use scale, NS = Nomophobia scale, ES = Possession incorporation in the extended self, SAt = Smartphone attachment, SAS = Smartphone addiction scale, SABAS = Smartphone application-based addiction scale PMPUQ = Problematic mobile phone use questionnaire, MTUAS = Media and technology usage and attitudes scale, SUQ-G = Smartphone use questionnaire (general), SUQ-A = Smartphone use questionnaire (absent minded), ObjT = Objective average daily screen-time, ObjP = Objective average daily number of pickups, ObjN = Objective average daily number of notifications.

* Correlation is significant at a 0.05 level (2-tailed).

** Correlation is significant at a 0.01 level (2-tailed).

In order to assess which estimates or measures performed the best when predicting behavior in general, we calculated the average correlation from all three objective measures (average time spent on their smartphone, average number of pickups, and average number of notifications), for each self-reported measure, and the three single estimates. From this, we note that the notification (NEst) [r = .33] and time (TEst) [r = .33] estimates had the highest average correlation with the three objective behavioral measures, closely followed by the Smartphone Usage Questionnaire – General (SUQ-G) [r = .31] and Smartphone Usage Questionnaire – Absent Minded (SUG-A) scales [r = .29] (Figure 1).

Figure 4.1. Average r value for each subjective measure across all three objective behavioral measures. Error bars illustrate standard error. Red indicates a single behavioral estimate. Dotted line represents mean correlation across all measures. Refer to Table 1 for abbreviations.

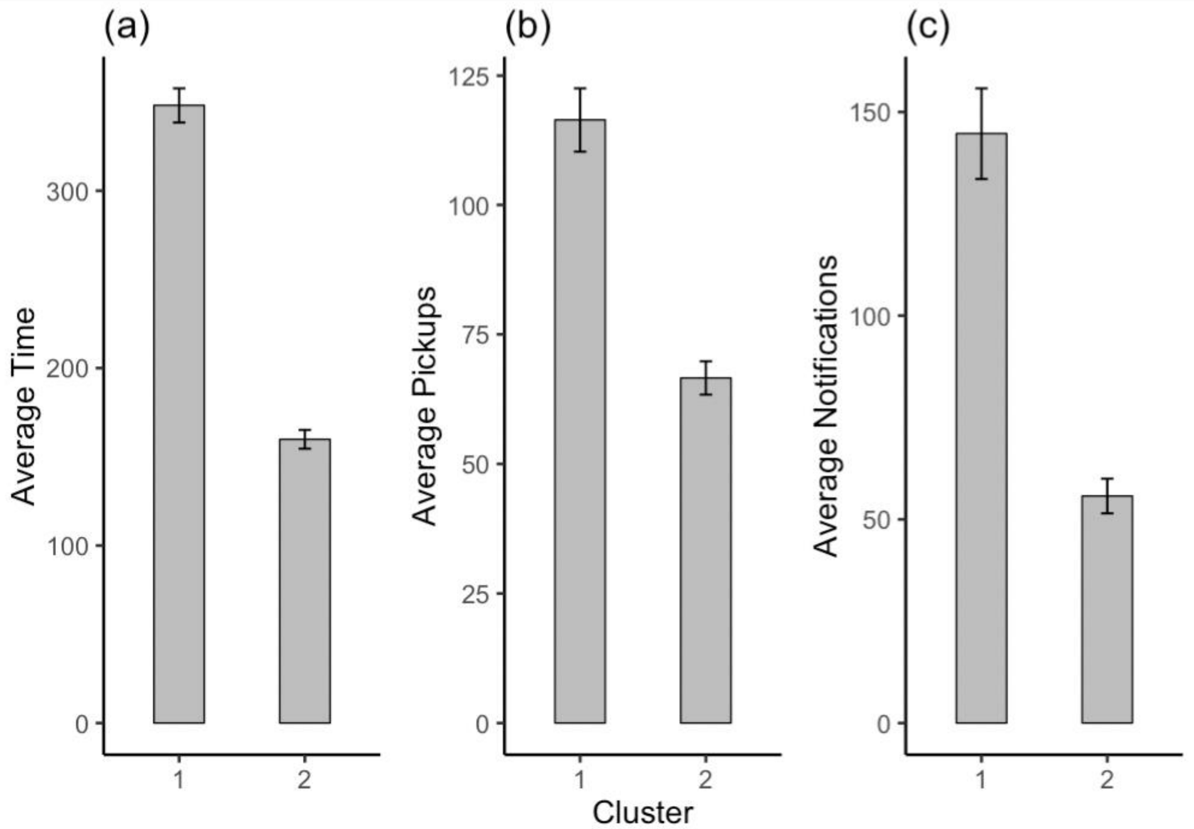


3.4 Cluster Analysis

Many conceptualizations of smartphone use focus on a binary classification whereby ‘addiction’ or ‘problematic’ usage are either present or absent. This is also important from a clinical standpoint as these scales are often referred to as having a (potential) diagnostic ability (Lin et al., 2016). Therefore, our final analysis considered if behavioral and self-report measures could classify identical participants. While several unsupervised methods can cluster participants, k-means is widely used in behavioral analytics (e.g., Arazy et al., 2017; Jackson, Østerlund, Maidel, Crowston, & Mugar, 2016; Wang, Brede, Ianni, & Mentzakis, 2018) because it can handle a variety of dataset sizes and produce straightforward outputs (Wu et al., 2008). The unsupervised nature of such an approach also removes any researcher bias.

Participants were clustered into two groups (high and low) twice with different input variables used for each classification. The first cluster analysis used *only* the three objective behavioral measures (time spent, notifications, and pickups). As expected, fewer participants scored highly in all three objective behavioral measurements. Figure 2 illustrates the means of high and low clusters for the objective behavioral measures.

Figure 4.2. Means of high (N = 92) (cluster 1) and low users (N = 146) (cluster 2) derived from objective data following a k-means cluster analysis. Error bars denote standard error.



A second cluster analysis used only self-reported scales (excluding single estimates) to make a similar distinction. Classifications for each participant were then compared. A large level of agreement between self-report and behavior would lead to identical participants being classified as high in both analyses. However, when comparing classifications between the two data-sets, only 52 of 92 (56.52%) participants identified as high users based on behavior, were also classified as high-users from self-report data.

As expected, the behavioral cluster analysis identified a large percentage (38.66%) of our sample as ʻhigh ʼusers. However, this may lack any meaningful specificity given that comp14aratively few participants are likely to demonstrate exceptionally high usage patterns (Wilcockson et al., 2018). As a result, research relying on self- report alone has considered non-binary approaches by adopting a three-cluster approach (Lepp, Li, Barkley, & Salehi-Esfahani, 2015). We therefore replicated our previous procedure with a three-cluster solution (k = 3), which separated users into low, medium, and high usage groups. Again, we compared clustering decisions derived from self-report and objective behavior. In this instance, the overlap of high users appearing in both clusters fell to 32.36% (10 out of 31). Here, we observe that moving away from a binary classification does not improve performance.

4 Discussion

To date, no systematic approach has attempted to behaviorally validate the growing number of psychometric instruments, which aim to capture technology related behaviors and experiences. Here, we demonstrate that smartphone related assessments are no better than single duration estimates when predicting subsequent behavior. However, as observed elsewhere, even single-item measurements fail to explain much of the variance associated with comparable behaviors (Boase & Ling, 2013). This has wide-ranging consequences for the vast number of studies that rely on these self-reported measures as a proxy measure of behavior.

Every psychometric scale correlated with at least one objective measure, but the strength of these relationships is far from convincing. Existing smartphone 'addiction' scales, for example, correlated poorly with the 'rapid checking' behaviors that one would associate with a behavioral addiction (Andrews et al., 2015; Rozgonjuk et al., 2018). As these scales struggle to capture simple behaviors, it remains questionable as to how they could effectively measure habitual, atypical, and more complex behavioral patterns. Further, combining multiple scales did not assist in the identification of participants with high usage patterns derived from behavior alone. As a consequence, our results have implications for studies that attempt to understand the impacts of smartphones and other screen-based technologies on health and wellbeing. These issues extend to research that has attempted to link a variety of individual differences (e.g., personality) with technology use (e.g., Butt & Phillips, 2008; Horwood & Anglim, 2018; Takao, Takahashi, & Kitamura, 2009). Errors of measurement here are so large that small effects detected in large-scale research involving estimates may be a component of statistical noise or a weak proxy for other psychological constructs (Ellis, 2019).

While the scales under investigation were developed in an effort to capture specific constructs (e.g., addiction or nomophobia), they are frequently used to quantify usage in the general population. This appears to be in direct conflict with a conceptual framework that problematizes usage without considering how typical these behaviors are within the general population. However, recent conceptualizations of usage perhaps hold some promise. The Smartphone Usage Questionnaires (SUQ) (Marty- Dugas & Ralph, 2018), provided the strongest correlations across the board. These consider everyday smartphone use in the context of attentional lapses and mind wandering instead of conceptualizing everyday behavior as 'addictive or 'problematic', which demonstrates the strength in focusing on cognition directly (e.g., attention to and distraction via technology) rather than addiction. These findings also align with recent theoretical models, which argue that technology use over time becomes habitual and more 'absent-minded '(Shaw et al., 2018). Indeed, a growing body of evidence now supports the notion that psychology should start to move away from a behavioral addictions framework when studying technology use (Panova & Carbonell, 2018).

Broadly speaking, technology usage assessments, which vary from television, to internet, online gaming, and more recently, smartphones, rely on extraordinarily similar scales or estimates – substituting device for device as required (Rosen et al., 2014). This similarity problem can also be considered within smartphone usage scales specifically. Despite being developed years apart and around different frameworks or conceptualizations of use (e.g., fear, attachment, or problematic use, etc.), they appear to, in many cases, measure almost identical constructs. The majority of smartphone usage scales by their very nature likely overlap with higher levels of anxiety and depression rather than smartphone usage, as the item's wording tends to be conceptually similar to that of depression and anxiety scales. One future study may wish to compare how these measures correlate with anxiety assessments and objective behavior. Our results suggest that the correlation would be far stronger with the former than the latter.

Given the complexities associated with studying the impact of technology on people and society, there is an urgent need for basic research to consider what this means for different individuals, devices, contexts, and in the case of smartphones, specific types of app usage (Jungselius & Weilenmann, 2018). The discipline may need to consider a paradigm shift, which would also help drive theoretical development and encourage a systematic shift away from the repetitive development of self-report assessments (Billieux, Maurage, Lopez-Fernandez, Kuss, & Griffiths, 2015). However, this may already be changing as Apple and Google are providing more of this data directly to all users, which provides a simple way to capture basic measures of objective behavior. We anticipate that this alone will lead to many other researchers making use of data derived from these screen time applications in the future. All this is not to suggest that there is no place for self-report or psychometric assessment in this domain of research at all. However, psychometric tools should be built around a concrete understanding of what (a) such measures can accurately assess and (b) what specific questions they can answer. For example, while functions of addiction can go beyond use (e.g., craving), the consumption of technology continues to be frequently referenced as a key metric by researchers in this domain (Dowling & Quirk, 2009). There are also certainly more specific behaviors, which might better map onto these psychometric scales, but research to date typically focuses on time spent on a device overall rather than specific sub-sets of behavior (Ellis et al., 2018). This has further implications for smartphone 'addiction' if it were to ever be included as part of the World Health Organization's ICD-11 (2018) alongside gaming disorder, as any diagnostic criteria will almost certainly have to focus on objective behavior, as well as thoughts, attitudes and feelings towards a technology (Lin et al., 2016).

4.1 Limitations

There are some limitations to note. First, while the behavioral measures utilized here are limited (e.g., this study uses daily tracking rather than finer grain temporal measurements based on hourly patterns of usage), we would argue that actually exploring interactions with technology directly provides a more suitable pathway moving forward. A second limitation concerns our specific use of Apple's Screen Time because this system allows participants to view their own data in real-time, which may partly explain why self-reported estimates correlated more favorably with objective behavioral measures. For example, self-reported pickups have previously not shown a relationship with objective behavior in a smaller sample (Andrews et al. (2015). However, the consistency of our results coupled with reminding participants to not look at their devices when providing estimates suggests that an alternative explanation is unlikely. A related issue may concern the omission of Android users, and previous research has suggested that behaviors and personalities differ between iPhone and Android platforms (Shaw, Ellis, Kendrick, Ziegler, & Wiseman, 2016). However, Andrews et al. (2015) reported an almost identical number of daily smartphone pickups (84.68) with a small number of Android users, demonstrating that regardless of operating systems, the average number of pickups reported in our sample remain remarkably similar. Perhaps more importantly, our findings echo earlier validation concerns albeit on a larger scale (Andrews et al., 2015; Elhai et al., 2018; Lin et al., 2015; Rozgonjuk et al., 2018; Wilcockson et al., 2018).

5 Conclusions

Here we attempted to validate smartphone usage scales against a handful of behavioral metrics. Our results suggest that the majority of these self-report smartphone assessments perform poorly when attempting to predict objective smartphone behaviors. Researchers should therefore be cautious when using these measures to link technology use with outcomes concerning health and psychological well-being. They also provide weak evidence to support the development of any diagnostic criteria (e.g., Lin et al., 2016; Tran, 2016). The issues highlighted here feed into a growing consensus that while psychology has acknowledged a problem with replication, the discipline also needs to address similar issues within measurement (Flake & Fried, 2019). Across psychological science, many self-reports remain insufficient for researchers who continue to make large claims, particularly those which pertain to the impact of technology on public health (Boyd & Pennebaker, 2017; Twenge, Joiner, Rogers, & Martin, 2017). We would encourage other researchers where possible, to complement these with objective measures of behavior in order to better understand the impact of technology on people and society more generally.

References

Abdi, H. & Williams, L. J. (2010). Principle component analysis. *WIREs Computational Statistics*, 2, 433-45

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond self-report: Tools to compare estimated and real-world smartphone use. *PLoS ONE, 10*(10), 1–9. https://doi.org/10.1371/journal.pone.0139004

Arazy, O., Lifshitz-Assaf, H., Nov, O., Daxenberger, J., Balestra, M., & Coye, C. (2017). On the "How" and "Why" of Emergent Role Behaviors in Wikipedia. In *In the proceedings of the Conference on Computer-Supported Cooperative Work Conference (CSCW)* (pp. 2039–2051).

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Belk, R. (1988). Possessions and the Extended Self. *Journal of Consumer Research*, *15*(2), 139–168. https://doi.org/10.1086/209154

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.

Bianchi, A., & Phillips, J. G. (2005). Psychological Predictors of Problem Mobile Phone Use. *CyberPsychology & Behavior*. https://doi.org/10.1089/cpb.2005.8.39

Billieux, J., Maurage, P., Lopez-Fernandez, O., Kuss, D. J., & Griffiths, M. D. (2015). Can Disordered Mobile Phone Use Be Considered a Behavioral Addiction? An Update on Current

Evidence and a Comprehensive Model for Future Research. *Current Addiction Reports*. https://doi.org/10.1007/s40429-015-0054-y

Billieux, J., Van Der Linden, M., & Rochat, L. (2008). The role of impulsivity in actual and problematic use of the mobile phone. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.1429

Boase, J., & Ling, R. (2013). Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*. https://doi.org/10.1111/jcc4.12021

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, *18,* 63–68. https://doi.org/10.1016/j.cobeha.2017.07.017

Brems, M., (2017). A One-Stop Shop for Principal Component Analysis. *Towards Data Science*. [Online]: https://towardsdatascience.com/a-one-stop-shop-for- principal-component-analysis-5582fb7e0a9c. [Accessed on]: 5/9/2019

Butt, S., & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2007.01.019 Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective

Chivers, T. (2018, December). Have we got social media all wrong? New scientist. 3207, 15.

Chivers, T. (2019, January). Stop scaremongering about kids spending time on their phones: Correlation does not equal causation when it comes to screen time. The New Statesman. Retrieved from: https://www.newstatesman.com/politics/uk-politics/2019/01/stop-scaremongering-about-kids-spending-time-their-phones

Scale Development. *Psychological Assessment*, *7*(3), 309–319.

Csibi, S., Demetrovics, Z., & Szabó, A. (2016). A Rövid Okostelefon Addikció Kérdoív (ROTAK) kidolgozása és pszichometriai validálása iskoláskorú gyermekekkel. TT - [Development and psychometric validation of the Brief Smartphone Addiction Scale (BSAS) with schoolchidren]. *Psychiatr Hung*.

Dolinski, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, *13*(2), e25025.

Doughty, M., Rowland, D., & Lawson, S. (2012). Who is on your sofa? TV Audience Communities and Second Screening Social Networks. In *10th European conference on Interactive tv and video*. https://doi.org/10.1145/2325616.2325635

Dowling, N. A., & Quirk, K. L. (2009). No Title. *CyberPsychology & Behavior*, *12*(1), 21–27. https://doi.org/10.1089/cpb.2008.0162

Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2016.08.030

Elhai, J. D., Levine, J. C., Dvorak, R. D., & Hall, B. J. (2016). Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2016.05.079

Elhai, J. D., Tiamiyu, M. F., Weeks, J. W., Levine, J. C., Picard, K. J., & Hall, B. J. (2018). Depression and emotion regulation predict objective smartphone use measured over one week. *Personality and Individual Differences*, *133*(21–28). https://doi.org/10.1016/j.paid.2017.04.051

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*, *97*, 60–66. https://doi.org/10.1016/j.chb.2019.03.006

Ellis, D. A., Kaye, L. K., Wilcockson, T. D. W., & Ryding, F. C. (2018). Digital Traces of Behaviour Within Addiction: Response to Griffiths (2017). *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-017-9855-7

Ellis, D. A., & Piwek, L. (2018). Failing to encourage physical activity with wearable technology: what next? *Journal of the Royal Society of Medicine*. https://doi.org/10.1177/0141076818788856

Flake, J. K., & Fried, E. I. (2019). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *PsyArXiv*. https://doi.org/10.31234/osf.io/hs7wm

Foerster, M., Roser, K., Schoeni, A., & Röösli, M. (2015). Problematic mobile phone use in adolescents: derivation of a short scale MPPUS-10. International Journal of Public Health, 60(2), 277–286. https://doi.org/10.1007/s00038-015-0660-4

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. Biometrika, 58(3), 453-467

Horwood, S., & Anglim, J. (2018). Personality and problematic smartphone use: A facet-level analysis using the Five Factor Model and HEXACO frameworks. *Computers in Human Behavior*, *85*, 349–359. https://doi.org/10.1016/j.chb.2018.04.013

Jaadi, Z., (2019). A step by step explanation of Principal Component Analysis. *Towards Data Science*. [Online]: https://towardsdatascience.com/a-step-by- step-explanation-of-principal-component-analysis-b836fb9c97e2. [Accessed on]: 5/9/2019

Jackson, C., Østerlund, C., Maidel, V., Crowston, K., & Mugar, G. (2016). Which Way Did They Go? Newcomer Movement through the Zooniverse. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16,* 623–634. https://doi.org/10.1145/2818048.2835197

Jones, S. L., Ferreira, D., Hosio, S., Goncalves, J., & Kostakos, V. (2015). Revisitation analysis of smartphone app use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 1197–1208). Osaka, Japan. https://doi.org/10.1145/2750858.2807542

Jungselius, B., & Weilenmann, A. (2018). Conceptualizing "Use" in Social Media Studies. In *Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18*. https://doi.org/10.1145/3217804.3217938

Kwon, M., Kim, D. J., Cho, H., & Yang, S. (2013). The smartphone addiction scale: Development and validation of a short version for adolescents. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0083558

Lepp, A., Li, J., Barkley, J. E., & Salehi-Esfahani, S. (2015). Exploring the relationships between college students 'cell phone use, personality and leisure.

Computers in Human Behavior. https://doi.org/10.1016/j.chb.2014.11.006

Lin, T. T. C., Chiang, Y.-H., & Jiang, Q. (2015). Sociable People Beware? Investigating Smartphone V ersus Nonsmartphone Dependency Symptoms Among Young Singaporeans. *Social Behavior and Personality: An*

International Journal. https://doi.org/10.2224/sbp.2015.43.7.1209

Lin, Y. H., Chiang, C. L., Lin, P. H., Chang, L. R., Ko, C. H., Lee, Y. H., & Lin, S. H. (2016). Proposed diagnostic criteria for Smartphone addiction. *PLoS ONE*.https://doi.org/10.1371/journal.pone.0163010

Marty-Dugas, J., & Ralph, B. C. W. (2018). The Relation Between Smartphone Use and Everyday Inattention. *Psychology of Consciousness: Theory, Research, and Practice*, *5*, 46–64. https://doi.org/10.1037/cns0000131

McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.52.1.81

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691612441215

Mou, D. (2016). Battling severe mental illnesses with smartphones: how patients 'smartphone data cna help improve clinical care. *MHealth*, *2*.

Oliver, E. (2010). The challenges in large-scale smartphone user studies. In *Proceedings of the 2nd ACM International Workshop on Hot Topics in Planet- scale Measurement - HotPlanet '10*. https://doi.org/10.1145/1834616.1834623

Panova, T., & Carbonell, X. (2018). Is smartphone addiction really an addiction? *Journal of Behavioral Addictions*. https://doi.org/10.1556/2006.7.2018.49

Parker, W. D., & Stumpf, H. (1998). A validation of the five-factor model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences*. https://doi.org/10.1016/S0191- 8869(98)00016-6

Piwek, L., Ellis, D. A., & Andrews, S. (2016). Can programming frameworks bring smartphones into the mainstream of psychological science? *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01252

Przybylski, A. K., & Weinstein, N. (2017). A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations Between Digital-Screen Use and the Mental Well-Being of Adolescents. Psychological Science, 28(2), 204–215. https://doi.org/10.1177/0956797616678438

Richardson, M., Hussain, Z., & Griffiths, M. D. (2018). Problematic smartphone use, nature connectedness, and anxiety. *Journal of Behavioral Addictions*. https://doi.org/10.1556/2006.7.2018.10

Rosen, L. D., Lim, A. F., Felt, J., Carrier, L. M., Cheever, N. A., Lara-Ruiz, J. M., ... Rokkum, J. (2014). Media and technology use predicts ill-being among children, preteens and teenagers independent of the negative health impacts of exercise and eating habits. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2014.01.036

Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rokkum, J. (2013). The Media and Technology Usage and Attitudes Scale: An empirical investigation. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2013.06.006

Rozgonjuk, D., Levine, J. C., Hall, B. J., & Elhai, J. D. (2018). The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2018.05.019

Shaw, H., Ellis, D. A., Kendrick, L.-R., Ziegler, F., & Wiseman, R. (2016). Predicting Smartphone Operating System from Personality and Individual Differences.

Cyberpsychology, Behavior, and Social Networking, *19*(12), 727–732. https://doi.org/10.1089/cyber.2016.0324

Shaw, H., Ellis, D. A., & Ziegler, F. V. (2018). The Technology Integration Model (TIM). Predicting the continued use of technology. *Computers in Human Behavior*, *83*, 204–214. https://doi.org/10.1016/j.chb.2018.02.001

Sivadas, E., & Venkatesh, R. (1995). An Examination of Individual and Object- Specific Influences on the Extended Self and its Relation to Attachment and Satisfaction. *Advances in Consumer Research*.

Takao, M., Takahashi, S., & Kitamura, M. (2009). Addictive Personality and Problematic Mobile Phone Use. *CyberPsychology & Behavior*. https://doi.org/10.1089/cpb.2009.0022

Tao, S., Wu, X., Zhang, S., Tong, S., Hao, J., & Tao, F. (2017). Association of alcohol use with problematic mobile phone use and depressive symptoms among college students in Anhui, China. *Journal of Public Health (Germany)*. https://doi.org/10.1007/s10389-016-0766-z

Tran, D. (2016). Classifying Nomophobia as Smart-Phone Addiction Disorder. *UC Merced Undergraduate Research Journal*. https://doi.org/10.1111/desc.12173

Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2017). Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among

U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. *Clinical Psychological Science*, *6*(1), 3–17. https://doi.org/10.1177/2167702617723376

UK Parliament. (2018). Impact of social media and screen-use on young people’s health inquiry launched. Retrieved August 24, 2018, from https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2017/social-media--young-peoples-health-inquiry-launch-17-19/

Wang, T., Brede, M., Ianni, A., & Mentzakis, E. (2018). Social interactions in online eating disorder communities: A network perspective. *PLoS ONE*, *13*(7), 1–17. https://doi.org/10.1371/journal.pone.0200800

Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining Typical Smartphone Usage: What Data Do We Need? *Cyberpsychology, Behavior, and Social Networking*, *21*(6), 395–399. https://doi.org/10.1089/cyber.2017.0652

Wolniewicz, C. A., Tiamiyu, M. F., Weeks, J. W., & Elhai, J. D. (2018). Problematic smartphone use and relations with negative affect, fear of missing out, and fear of negative and positive evaluation. *Psychiatry Research*. https://doi.org/10.1016/j.psychres.2017.09.058

World Health Organization (WHO). (2018). ICD-11 International Classification of Diseases 11th Revision. Retrieved March 3, 2019, from https://icd.who.int/

Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). *Top 10 algorithms in data mining*. *Knowledge and Information Systems* (Vol. 14). https://doi.org/10.1007/s10115-007-0114-2

Yildirim, C., & Correia, A. P. (2015). Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2015.02.059

Zhao, S., Ramos, J., Tao, J., Jiang, Z., Li, S., Wu, Z., ... Dey, A. K. (2016). Discovering Different Kinds of Smartphone Users Through Their Application Usage Behaviors. In

Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 498–509). Heidleberg, Germany. https://doi.org/10.1145/2971648.2971696

Chapter VI

Quantifying smartphone ‘use’: Choice of measurement impacts the relationships between ‘usage’ and health

Chapter Publication status: Published

Shaw, H., Ellis, D., Geyer, K., Davidson, B., Ziegler, F., & Smith, A. (2020). Quantifying smartphone “use”: Choice of measurement impacts relationships between “usage” and health. *Technology, Mind, and Behavior*, *1*(2).

Specific aim

3. Deploy a psych app that I developed within empirical research

My contribution

I developed a specific version of an early psych app in order to conduct this research. Altered the design of the psych app in response to the results back from the pilot study. For study 2, I came up with the primary rational and specified what data should be collected. I edited drafts of the article and added additional analysis such as controlling for multiple comparisons.

The previous chapter unequivocally demonstrated the limitations of self-report methods when compared to objective methods of measuring behaviour. There were negligible correlations between self-report questionnaires on smartphone usage and actual usage. This finding represents a substantial challenge to previous claims that smartphone 'addiction' is even a valuable construct. While there is more to an addiction than just behaviour, 'validated' questionnaires all performed worse than simply asking participants in a single question - how much do you use your smartphone. However, the study reported in *Chapter V* only used third party software which is not transparent in how it operates. The accuracy of this system is difficult to assess. Therefore, we used a previously validated psych app (*Chapter IV*) to see if the results can be replicated and extended.

This chapter concludes that some psych apps can overcome limitations with applied methods in social psychology. Subjective methods such as questionnaires have been historically popular, especially within social psychology. Given the problems associated with these methods, questions emerge. Are the assertions of studies that rely upon questionnaires flawed?

Specifically, are smartphones at a general level having a negative impact on health and wellbeing.

In this chapter, we report two studies. The first study involved 46 participants, who were extensively reviewed for their smartphone usage by using Usage Logger (outlined in *Chapter IV*) and also their physical health and mental health. We analysed if there was a relationship between the amount that the smartphone was being used and the participants' physical and mental health. Study 2, again used a Apple Screen Time (Apple, 2019) that returned lower resolution data, but allowed for more participants (199) to be involved.

This final contribution centred on the issues around how spurious conclusions can be drawn from employing flawed measures. When psychology employs inaccurate methods then the resulting theory, conclusions and policy implications become distorted. This published paper, suggested that moral panics regarding smartphone 'addiction' may be built around limited methodologies. Consequently, when appropriate methods are employed no public cause for alarm emerges when it comes to high-level measures of general technology use.

Abstract

Problematic smartphone use scales and estimates of use dominate research that considers the impact of smartphones on people and society. However, issues with conceptualisation and subsequent measurement may obscure any genuine associations between technology use and mental health. Here, we considered whether different ways of measuring 'smartphone use', notably through problematic smartphone usage (PSU) scales, subjective estimates, and objective logs, leads to contrasting associations with both mental and physical health. Across two samples including iPhone (*n*=199) and Android (*n*=46) users, we observed that measuring smartphone interactions with PSU scales produced larger effect sizes with mental health than subjective estimates or objective logs. Notably, the size of the relationship was fourfold in study 1, and over twice as large in study 2 when employing a smartphone addiction scale in comparison to objective measures. Additionally, findings showed positive relationships between average daily steps and average daily walking and running distance with objective daily pickups. This questions whether all smartphone behaviors should be considered sedentary. To conclude, addressing people's concerns about their usage is likely to have greater mental health benefits than reducing their overall device use, and should not be a priority for public health interventions at this time.

Keywords: Smartphones, Technology, Mental Health, Sedentary Behaviors, Screen Time

5.1 Introduction

Smartphones are devices primarily used for connecting people in personal and occupational settings. Yet, much understanding of the relationship between smartphone use and health has been dominated by research which focuses on the ‘negative consequences’ of smartphone use and screen time with a strong focus on mental health (Elhai, et al., 2017), but also sedentary behaviour and physical activity (Zagalaz-Sánchez, et al., 2019). Coined ‘problematic smartphone use’ (Elhai et al., 2017), these perceived undesirable side-effects of use are also mirrored in public discourse (Genc, 2014; Yang, Asbury, & Griffiths, 2019). However, there is now some acknowledgement that existing research linking any screen time behaviours to health outcomes are weak, which makes it difficult for governments to make policy decisions (Science-and-Technology-Commitee, 2019). Specifically, research needs to address issues with measurement (Ellis, 2019), theory (Orben, 2018; Shaw, Ellis, & Ziegler, 2018), analysis choices (Orben & Przybylski, 2019), and prioritise high-quality designs to better understand genuine benefits or harms (Coyne, et al., 2019; Heffer, et al., 2019). In this paper, we specifically investigated whether the relationship between smartphone use and health changes noticeably as a result of measurement choices. This may in part explain the lack of a coherent academic position regarding the impact of smartphone use on wellbeing, whilst also having implications for those curating interventions.

Existing survey research has linked increased smartphone screen time to lower psychological wellbeing (Twenge, Martin, & Campbell, 2018), yet research using objective logs suggests the opposite (Katevas, Arapakis, & Pielot, 2018). Therefore, the diverse range of methods used to measure technology use across research designs could explain conflicting results, ranging from duration and frequency estimates, to psychometric scales, and objective logs (Elhai, et al., 2017; Harwood, et al., 2014; Rozgonjuk, et al., 2018; Vahedi & Saiphoo, 2018). One review (*n*=41) observed across disciplines that it was popular to measure smartphone use by asking participants to estimate their usage frequency (40% of papers), estimates of durations of use (27% of papers) or through other forms of self-report measures (9% of papers) (Boase & Ling, 2013). In recent years, concerns regarding 'overuse' have led to an abundance of usage scales being created to measure new constructs, including: 'addiction', 'nomophobia', and 'problematic use' (Ellis, 2019; Thomée, 2018). Others have noted an apparent absence of measuring smartphone use with direct and objective data available directly from devices themselves (Ellis, et al., 2019; Twenge, 2019). Whilst there is no consensus regarding how smartphone usage or screen time should be captured, many papers make claims suggesting that usage is the primacy variable of interest (Ellis, 2019).

In line with this, research has attributed greater smartphone use to increased sedentary behaviours (Lepp, et al., 2013; Zagalaz-Sánchez et al., 2019). Accordingly, people report that 87% of all phone use occurs whilst seated (Barkley & Lepp, 2016) and similarily, 90.9% of users report that they typically are sitting when using their smartphone (Xiang et al., 2020). Thus theoretically, it has been proposed that increased smartphone use lowers energy expendenture due to sedentary behaviours, and it is this mechanism which links increased use to greater body fat and obesity (Hamilton, Hamilton, & Zderic, 2007; Kim, Kim, & Jee, 2015). However, whilst 14 out of 9 articles in a recent systematic review showed a negative relationship between smartphone use and physical activity, none of the articles measured smartphone use objectively via logs from the device itself. Instead, people self-reported the duration and frequency of their smartphone behaviours, which is widely documented to only have moderate correlations with actual usage (Andrews et al., 2015; Boase & Ling, 2013; Parslow, Hepworth, & McKinney, 2003; Ellis et al., 2019; Kobayashi & Boase, 2012; Lee, et al., 2017; Vrijheid et al., 2006). Therefore, the research linking physical activity or sedentary behaviours to smartphone use is both scarce and yet to be examined precisely using objective smartphone behaviours.

When understanding mental health relationships, more nuanced approaches suggest that the way people appraise their smartphone use can be linked to wellbeing. Therefore, health outcomes might not necessarily be due to the physical use of the device itself. For example, a recent study found no robust evidence linking objective use of social apps to momentary wellbeing (Johannes et al., 2019). However, they did find that the more positively people felt about technology-mediated interactions in the past half hour, the better they felt in the current moment (Johannes et al., 2019). Furthermore, when assessing email use in occupational settings, stress occurs when a person perceives their usage to be in greater or lower amounts than is desired (Stich, et al., 2019). This is in line with the cognitive behavioural approach which suggests that our thoughts and beliefs can influence our emotions and behaviours, and when we have distorted perceptions of our experiences, this promotes negative mood states (Beck, 1967). Addressing negative thought cycles during cognitive behavioural therapy has been identified as an effective treatment for depression and generalized anxiety disorder, supporting the notion that these psychopathologies involve cognitive appraisals (Butler, et al., 2006). Thus, it is plausable that the way people percieve their smartphone usage behaviours (e.g. a belief that their smartphone use is excessive) may drive the relationships with mental health, seperately from the usage itself.

However, more often than not, researchers claim to be measuring smartphone usage, when instead they are measuring people's appraisal of use. For example, defining or measuring problematic smartphone use (PSU) in relation to 'overuse' or 'excessive use' is prevalent in many recent articles (Elhai & Contractor, 2018; Elhai, et al, 2020; Kim, 2017; Yang et al., 2019). This is arguably because it has foundations in the addiction framework whereby tolerance is a component (the need to increase use over time to get the same 'fix'), and also because recent conceptualisation of PSU discuss the similarities between PSU and addiction (Billieux, Maurage, et al, 2015; Elhai et al., 2017; Kim, 2017). Consequently, it is not surprising to find questions such as *"Using my smartphone longer than I had intended"*, and *"Having tried time and again to shorten my smartphone use time but failing all the time"* in problematic usage scales (Kwon et al. 2013). However, agreeing with these statements only shows that a person is negatively appraising their smartphone use, and is not a measure of frequency in itself. Correspondingly, research correlating problematic usage scales with objective smartphone usage show small to medium effect sizes (Ellis et al. 2019), and factor analysis research shows that PSU scores weakly load on factors representing actual usage (Davidson, Shaw, & Ellis, 2020). This evidence demonstrates how people's appraisals of their smartphone use and actual usage should be understood seperately, which is not currently the case.

In light of this unclear conceptualisation, it is important to distinguish between PSU as a psychological construct which appraises use, and smartphone usage as a behavioural variable, because it has implications for theory and treatment. For example, pop-up notifications containing usage statistics is one proposed PSU intervention that aims to reduce usage (Loid, Täht, & Rozgonjuk, 2020) and has recently been implemented by Apple in contemporary iPhone operating systems. However, if studies do not measure usage, then it might be incorrect and ineffective to promote this as a treatment option. For example, in a study linking greater fat mass, lower muscle mass, and lower daily step count to higher smartphone addiction scores, it was proposed that more time on smartphones equated to less physical activity (Kim, Kim, & Jee, 2015). However, usage was not measured in this study. To address this limitation, and the current research gap when investigating physical health outcomes, the present study measures body composition, body mass index and daily step count, as indicators of physical activity, to start understanding if this has relationships with objective smartphone use.

Additionally, it is important to distinguish between PSU and smartphone usage, as the conclusions people make regarding the relationship between smartphone use and mental health appears to depend on the measurement used (Vahedi & Saiphoo, 2018). In a systematic review of 23 studies, anxiety, stress, and depression were consistently linked with scores on problematic smartphone use scales (Elhai et al., 2017). However, when researchers measure smartphone usage instead of collecting PSU scores the relationships seem to diminish. For example, when using an online survey, Harwood et al. (2014) found that self-reported frequency of use was not associated with depression, anxiety or stress measures. Moreover, when using objective logs, Rozgonjuk et al., (2018) found that screen time minutes over a weeklong period was not related to depression and anxiety. In another study, intense objective smartphone use did not predict negative wellbeing (Katevas et al., 2018). Finally, a meta-analysis of 37 studies found that when measuring smartphone use through scales such as 'The Smartphone Addiction Scale' (Kwon et al., 2013), associations between stress and anxiety were stronger when compared to self-reported frequencies of use (Vahedi & Saiphoo, 2018). Consequently, measuring associations between health and smartphone use in different ways appears to generate radically different results across three documented measures: subjective estimates, objective logs, and psychometric scales. If academics were to interpret all these studies as measuring smartphone use, instead of appraisals in some cases, then the collective message would be conflicting and misleading. Of importance to the present study, conclusions appear to be an artefact of the measurement used.

Consequently, this paper aims to understand this issue by collecting subjective estimates, objective logs, and psychometric scales from the same participants, to systematically assess the above pattern. We therefore asked the question *"Do scores on a problematic use scale have stronger relationships with health than measures of usage within the same users?* Furthermore, we examined the notion of 'overuse', separately from people's perceptions of their use, by exploring the size of the linear relationship between smartphone estimates, logs, and health. Therefore, we also ask *"Does increased smartphone use relate to lower mental* *wellbeing and poorer physical health?"* This was to inform policy regarding setting screen time limits and evaluate treatments which advocate reducing overall usage. These ideas were first investigated during exploratory analysis of 46 adults who completed all three measurements, alongside an assessment of their body composition and anxiety, depression and stress symptomology. The results were then used to create hypotheses regarding the influence of measurement on effect sizes. A second study was then conducted as an upscaled replication of the first study but with increased statistical power, to allow for greater confidence in the conclusions being made.

5.2 Study 1

5.2.1 Methods

5.2.1.1 Participants

The sample consisted of 46 [12 male] participants that were staff and students from the University of Lincoln, UK. Power calculations determined that a total sample size of 44 was adequate to investigate two-tailed medium-to-large effect sizes ($r > .4$) with a power of .8 when $\alpha = .05$. Age was skewed, as we tested predominately younger adults [$M = 23.54$, $SD = 8.25$]. All participants were Android smartphone users and stated they exercised less than 10 hours per week. There was originally 80 participants recruited but 36 of the participants had issues logging with the data logging on the psych app. This is a common occurrence with android psych apps (Saeb et al., 2015).

The study was advertised around the University campus using posters, leaflets, subject pool systems, and social media channels during term time and during public engagement events. Therefore, the sample consisted of those who emailed the researcher in response to these advertisements. Participants were told they would receive a graph of their phone use and a printout of their health analysis as incentives to take part. Those recruited through subject pool systems received course credit in compensation for their time.

5.2.1.2 Measures

Study 1 collected numerous variables to explore the relationships between individual differences and objective smartphone use. For brevity, the focus of this manuscript is to describe the body composition and mental health relationships with general smartphone use. Therefore, only the variables and data collection procedures related to this aim are described here. For further information on all the variables collected see supplementary materials.

Objective Smartphone Use

Objective smartphone data was collected using an app developed specifically for the project called Activity Logger (Geyer, 2018). This ran on Android devices and collected data to the resolution of one second. Activity logger was set up to listen to three events: the phone being turned on, the screen being activated, and the screen being turned off. Background operations then took this information, retrieved the current time stamp, and stored this in internal memory. This data file was then exported via the app and contained a list of records where a UNIX time stamp was paired with an event stating whether the screen was turning "ON" or "OFF". Source code for the app is available to download (https://osf.io/a4p78/). Data was held in the apps' specified internal memory on the participants phone and would be automatically deleted when the app was deleted.

Estimates of Smartphone Use

To measure people's estimates of their daily smartphone screen time, participants were asked one question: "*Think back to days 2 - 8 of the study. On average, how many hours a day did you spend on your smartphone?".* Participants responded in hours and minutes. To measure people's estimates of how many times a day they 'picked up' their device, participants were asked: "*Think back to days 2 - 8 of the study. On average, how many individual times did you use your smartphone a day? Think of these as individual pick-ups*."

Problematic Smartphone Use

Smartphone addiction was measured using the Smartphone Addiction Scale (SAS), which contained 33 items (Kwon et al., 2013). Participants rated the extent to which they agreed to several statements, for example "*Feeling pleasant or excited while using a smartphone*". Participants responded on a six-point Likert-Scale ranging from "Strongly Agree" (1) and "Strongly Disagree" (6). Higher scores indicated greater addiction risk. This scale was chosen because it is popular and widely cited and it correlates highly with other smartphone psychometric scales (Ellis et al. 2019; Thomée 2018). Additionally, it has been shown to be the prime example of a PSU scale when compared to other scales, represented by its high loadings on a PSU factor (Davidson, Shaw & Ellis, 2020). This measure was impart as utilized to see if the findings of a disconnect between objective use of smartphones and scores on questionnaires attempted to review smartphone use could be replicated in Android users (Ellis, Davidson, Shaw, & Geyer, 2019).

Anxiety

Symptoms of anxiety were measured using the GAD-7 (Spitzer, et al. 2006) and included 7 items. Participants were asked "*how often in the last two weeks have you been bothered by*…" and responded on a four-point scale whereby 0 = "Not at all" and 3 = "Several Days". Using >10 as a cut-off point, the GAD-7 has been shown to have 89% sensitivity and 82% specificity with a diagnosis of general anxiety disorder (Kroenke et al., 2007).

Depression

Severity of depression was measured using the PHQ-9 (Kroenke, et al. 2001). Each of the nine questions related to a criterion mentioned in the DSM-IV for depression. Participants were asked "*how often in the last two weeks have you been bothered by…*" and responded on a four-point scale whereby 0 = "Not at all" and 3 = "Several Days". Using >10 as a cut-off point, the PHQ-9 has been shown to have 88% sensitivity and 88% specificity with a diagnosis of major depression (Kroenke et al., 2001).

Perceived Stress

The Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983) had 14 items which measured 'the degree to which situations in one's life are appraised as stressful'. Participants responded how often they felt a certain way on a 5-point Likert scale whereby 0 = "Never" and 4 = "Very Often". Participants were asked questions such as "*In the last month, how often have you felt that you were on top of things?*". Higher scores indicated greater perceived stress.

Objective Health Measures

Height was measured using a meter stick, with age and gender captured via self-report questions. This data was inputted as controls in subsequent bioimpedance analysis. Body composition was measured using the eight electrode Tanita MC-780MA body composition monitor. This provided an estimate of a person's body fat percentage, body mass index, and skeletal muscle mass percentage, using bioelectrical impedance measures. Bioelectrical impedance assessment using the Tanita MC-780MA was a good alternate to Magnetic Resonance Imaging and Dual Energy X ray absorptiometry (DEXA) which are costly, and time-consuming (Verney, et al., 2015). Notably, the Tanita MC-780MA produces body fat assessments which highly correlate with DEXA assessment ($r = .852$) providing concurrent validity (Verney et al., 2015).

5.2.2 Procedure

The study lasted nine days. On day one, a lab session provided participants with study information, including example data, followed by a consent form and an online questionnaire. Participants answered questions, including date of birth, gender, and a few psychometric scales beyond the scope of this manuscript (see supplementary materials). Once completed, participants were guided through the installation of the activity logger, and the researchers documented the smartphone brand and operating system. All screen savers were set to turn off after 30 seconds, and the application was 'white listed' in the smartphones' battery settings, ensuring that the phone would not 'force quit' some of the applications functions during low battery or processing power. Participants were then asked to keep their phone switched on for the duration of the study, and to keep the application running in the background. Whilst the application should re-start independently, as a precaution, if a participant's phone was switched off or had depleted battery during the week, participants were instructed to re-open the application once the phone had restarted. Participants were then provided with information detailing how to prepare for the body composition assessment on day nine. To control for factors influencing body composition results, participants were asked to refrain from intense exercise and alcohol up to 12 hours prior to the assessment, to keep hydrated, to book a time in the afternoon that was three hours after lunch, and to go to the toilet before the session.

Participants were then asked to use their phone as normal, and to carry on with their everyday activities across days two-eight of the study. This ensured that seven full days' worth of smartphone data was collected for each participant. On day nine, they returned to the lab and upon arrival, emailed the data from the app to the researcher. Next, participants completed a questionnaire containing the stress, anxiety, depression, smartphone addiction scales, and a few other measures beyond the scope of this manuscript (see supplementary materials). They were then asked to provide a daily average estimate of how much they picked up their phone, and the amount of time they spent on their phone across days 2-8.

Height was measured for the bioimpedance assessment. Participants were instructed to remove any jewellery, items in pockets and metal accessories, and were then asked stand bare foot on the Tanita MC-780MA body composition monitor while holding the hand electrodes by either side of their body, without touching their legs. A 0.5kg clothing allowance was inputted into the Tanita software if participants were wearing light clothing (gym gear), and a 1kg clothing allowance was inputted for heavy clothing (jumpers, jeans). Upon completion, participants were given a printout of their body composition, a graph of their application use, and of their screen time across the week. Finally, participants were debriefed and thanked for their time.

All procedures received ethical clearance by the School of Psychology Research Ethics Committee at the University of Lincoln and complied with British Psychological Society Guidelines (British Psychological Society, 2018). In the debrief, participants were told that the study would not offer any clinical diagnosis of any disorders and were provided with information about charities and services if they needed further support. The study also underwent a data protection plan. Participants had full control of their data as phone logs were stored solely on their devices and could be deleted by the participant at any point during the study by simply uninstalling the application.

5.2.3 Results

5.2.3.1 *Data processing*

Data and analysis scripts for study 1 can be found here (https://osf.io/a4p78/). The median daily hours-of-use was calculated across days two-eight for each person to remove influence of extreme "Screen On" events that occurred if the phone battery depleted and the app did not log a 'Screen Off" event. Daily pickups (frequency of use) were averaged across days two-eight, in accordance with recent work (see Ellis, et al, 2019). For the smartphone addiction scale, GAD-7, and PHQ-9, the responses were summed to create a total score for each scale. Specific questions within the perceived stress scale required reverse coding, and then an overall sum was created per person.

5.2.3.1 ***Exploratory Analysis***

When collating all 46 participants' data together, smartphone use was highly skewed, as 54.44% of uses were under 30 seconds in duration, and 43.54% of uses were under 15 seconds in duration. Due to this skew, we followed Bishara and Hittner (2017) recommendations and conducted Spearman Rank order correlations with Fieller, Hartley and Pearsons (1957) variance when calculating 95 % confidence intervals as these are robust against non-normality. To explore how differences in smartphone measurement may influence the size of the relationships found with health, Spearman correlations were conducted between all the health and smartphone variables (see Table 2). Notably anxiety, depression, and, stress had significant positive correlations with smartphone addiction scores (all p's <.01), which did not occur with any other smartphone measure.

Multiple comparisons were also controlled for via the Benjamini-Hockberg procedure (Benjamini, & Hockberg, 1995) as Bonferonni corrections would have been excessively conservative and risked the type II errors (Field, Miles, & Field, 2012). A false discovery rate of .1 was used as the study is exploratory and generally this is an excepted rate for exploratory studies (McDonald, 2014). Meaning we would expect that 10% of all statistical test would falsely report a significant difference.

Smartphone addiction scores consistently had r_S effect sizes larger than .39 with mental health, whereby estimates and objective variables were much lower (all r_S<.2) (see Fig. 3).

Table 5.1: Descriptive statistics for study 1 variables.

Health Measures	Mean	SD	α	Smartphone Measures	Mean	SD	α
Body Mass Index	24.84	5.86		Median Daily Screen Time (hrs)	3.74	1.60	.90
Body Fat %	26.97	8.86		Average Daily Pickups	133.18	63.52	.93
Skeletal Muscle Mass %	41.35	6.40		Daily Screen Time Estimate (hrs)	5.08	3.36	
Anxiety	6.13	5.56	.92	Daily Pickups Estimate	48.74	39.96	
Depression	6.57	5.25	.85	Smartphone Addiction Scale	90.09	21.20	.90
Stress	24.61	8.42	.87				

Table 5.2. Results of the Spearman correlations between smartphone and health variables from study 1.

Health Variable	Smartphone Addiction		Screen Time Estimate		Pickups Estimate		Median Daily Screen Time		Average Daily Pickups	
	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*
Physical Health										
Body Mass Index	-.25	-0.51, 0.05	-.10	-0.39, 0.21	-.14	-0.42, 0.16	-.32*[n]	-0.57, -0.03	-.39**	-0.62, -0.11
Body Fat %	.09	-0.21, 0.38	.18	-0.13, 0.45	-.01	-0.31, 0.29	-.01	-0.30, 0.29	-.12	-0.40, -0.18
Skeletal Muscle Mass %	-.06	-0.35, 0.24	-.14	-0.42, 0.17	.05	-0.25, 0.35	.06	-0.24, 0.35	.19	-0.11, 0.47
Mental Health										
Anxiety	.44**	0.17, 0.66	.11	-0.19, 0.40	.05	-0.25, 0.34	-.00	-0.30, 0.30	.11	-0.20, 0.39

Depression	.39**	0.11, 0.62	.19	-0.11, 0.47	-.05	-0.35, 0.25	.05	-0.25, 0.34	.08	-0.23, 0.37
Stress	.53***	0.27, 0.71	.18	-0.13, 0.45	.03	-0.27, 0.32	.00	-0.30, 0.30	.03	-0.27, 0.32

Notes: * significant at $p < .05$, ** significant at $p < .01$, *** significant at $p < .001$, [n] not significant after controlling for multiple comparisons

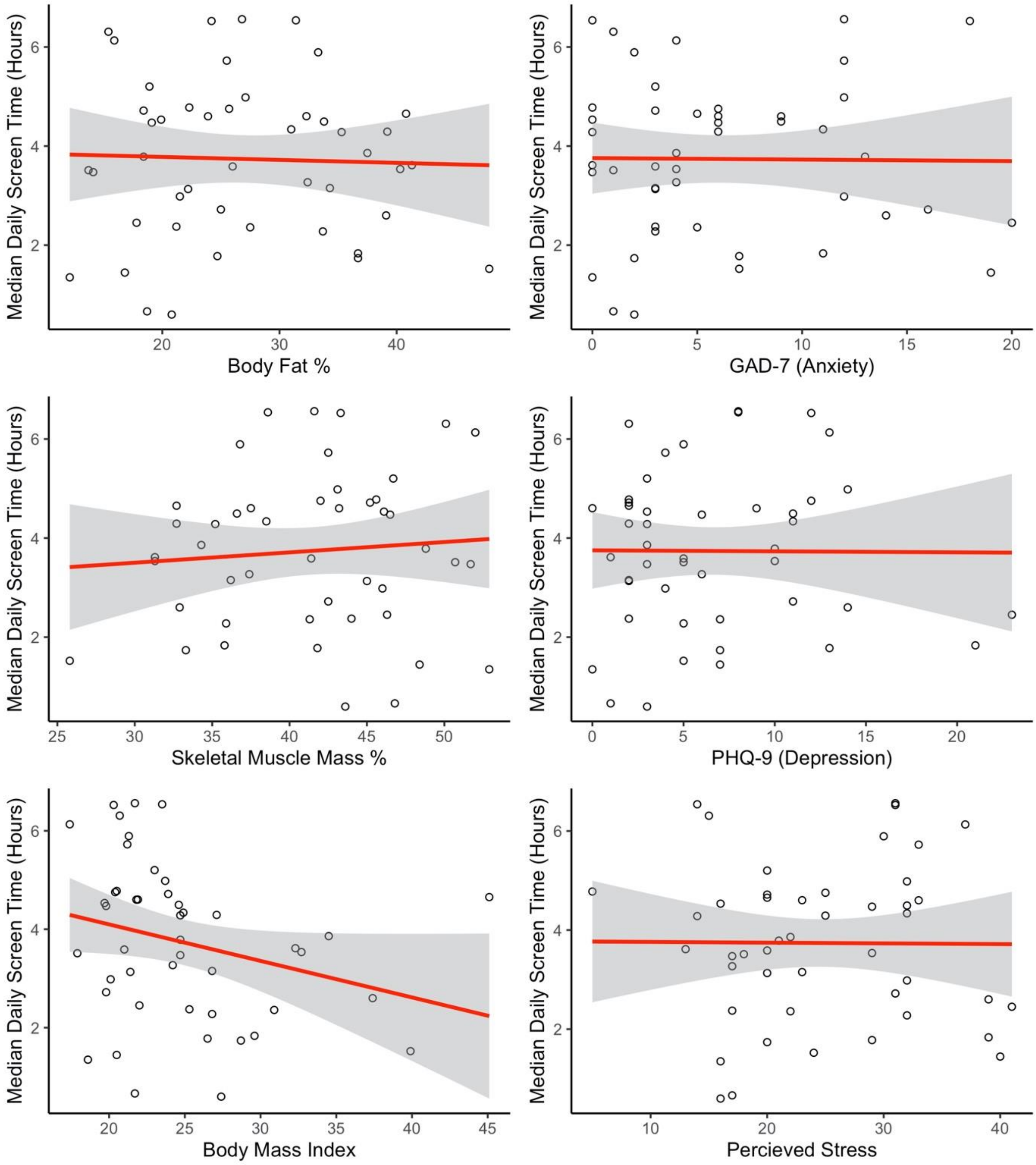


Figure 5.1. Scatter plots showing the linear relationships between median daily screen time (Hours) with six health variables; body fat percentage, skeletal muscle mass percentage, body mass index, anxiety, depression and stress. The red regression line represents the linear relationship between the two, and the surrounding grey area represents the 95% confidence interval.

5.3 Discussion

In study 1, smartphone addiction was found to positively correlate with anxiety, depression, and stress measures. Pertinently, effect sizes quadrupled when measuring smartphone usage with a problematic usage scale in comparison to objective screen time and pickup measures. Therefore, even within the same participants, a researcher could make different conclusions dependent on the measurement used, especially if confounding the construct problematic smartphone use with usage. In line with prior work, people's appraisals of their smartphone usage had stronger relationships with mental health than self-reported frequencies of use (Vahedi & Saiphoo, 2018) or objective logs (Rozgonjuk et al. 2018). This suggests people's worries regarding their smartphone use is more pertinent to mental health symptomology than the actual use itself.

Interestingly, we found that BMI reduced as daily screen time and pickups increased with significant effect. However, this finding was not deemed as reliable when the number of statistical comparisons were controlled for and the relationship was in the same direction, but much smaller when measuring body fat percentage. Our interpretation is that the study did not demonstrate evidence of a link between daily smartphone screen time and pickups on physical health. However, we marked these findings as tentative until they could be replicated on a larger sample. This was achieved in study 2, whereby we collected the same mental health and smartphone measures as study 1. We also re-assessed BMI via self-reports and took advantage of retrospective data collected on the user's device, including daily logs of steps, and daily logs of walking and running distance. Taking inspiration from study 1 findings, we predicted that effect sizes of $r_s > .3$ would be found when comparing mental health relationships with problematic usage scales, and that lower effect sizes of $r_s < .2$ would be found when examining estimates of use and objective logs.

5.3 Study 2

5.3.1 Methods

5.3.1.1 Participants

199 [137 women] participants, were recruited via Prolific Academic, from a subject pool of 24,117 iPhone owners. This pool contained predominately citizens from the United Kingdom and the United States. Participants had a mean age of 30.18 [SD = 9.46] and were paid £1.25 for their time. 42.71% of the sample were overweight or obese, and the average BMI across all participants was slightly higher than the recommended range [M = 25.17, SD = 5.38]. This was to be expected in a representative sample, as 52% of people have a BMI over 25 world-wide (WHO, 2018). A priori power calculation was performed which showed during two-tailed analysis a sample size of 192 participants was enough to detect small effect sizes of $r_S \geq .2$ with a power of .8 when $\alpha = .05$.

5.3.2.1 Measures and Procedure

Once clicking the link to access the online questionnaire, participants were presented with study information and a digital consent form. If participants agreed to take part, they were then asked; "*Please estimate how many hours and minutes you spend on your phone each day*" and answered in hours and minutes. In addition, participants were asked: "*Please estimate how many times a day you pick up and use your phone*". After, smartphone addiction, anxiety, depression, and stress were then measured using the same scales as study 1.

Objective smartphone usage data was retrieved by utilising the Apple Screen Time feature that resides in modern iPhones. We used the same methodology as Ellis et al. (2019)

and extracted data retrospectively from the past 7 days. In short, participants were prompted to find the 'Screen Time' graph and the 'Pickups' graph in Apple Screen Time settings and record for each day the number of pickups and screen time (in hours and minutes). For more details, see Ellis et al. (2019; or Chapter *IV*).

After obtaining objective smartphone use data, the questionnaire asked people to input their health data. The Apple Health App automatically tracks users' steps, walking, and running distances. This historic data is accessible on a user's iPhone for the entire time they have owned their iPhone. When clicking on the 'Today' tab, participants had access to a calendar where they could view their activity for any past day. Daily steps were collected by asking participants to click on the calendar pages for dates in the past week and enter for each day the number of steps displayed. Daily walking and running distances were collected by asking people to click on the calendar pages for dates in the past week and report the documented distance in either kilometres or miles. Participants were also asked if they owned a fitness tracker or a smartwatch and specified whether this device was synced to the Apple Health App. Lastly, participants were asked to report their age, gender, weight and height. They were given the option to answer in either metric (meters and centimetres / kilograms) or imperial measures (feet and inches / stones and pounds). At the end of the questionnaire, participants were debriefed, thanked for their time, and were then re-directed back to the prolific academic website.

All procedures received ethical clearance by the School of Psychology Research Ethics Committee at the University of Lincoln and complied with British Psychological Society ethical guidelines for internet mediated research (Hewson et al., 2013). Akin to study 1, the debrief provided websites where participants could get guidance on their mental health and were told details of 24-hour support lines. Participants could withdraw at any time before, during or up to two weeks after they completed the study by emailing the researcher.

5.3.3 Results

5.3.3.1 *Data removal*

Data and analysis scripts for study 2 can be found here (https://osf.io/a4p78/). The survey received 263 respondents. However, this became 207 after removing those who did not have iOS12 installed, did not have an iPhone 5 or later, did not have seven days of screen time data on their smartphone, or did not complete the survey or fill in the health questions. A further person was removed after being identified as an outlier when plotting data; they further had weight and BMI values more than three standard deviations from the mean. Finally, seven people were removed due to input errors (typos) in their health data. This left 199 participants in the following analysis.

5.3.3.2 *Data coding and processes*

Average daily screen time and average daily pickups scores were computed per person by taking the daily amount of screen time/pickups from the first six days and then calculating the mean. Six rather than seven days were used to compute this mean, as data from the seventh day did not represent a full day. Raw estimated number of daily pickups and estimated average daily screen time (in hours) were used in the analysis. Smartphone addiction and anxiety, stress and depression scales were all scored in the same way as study 1.

The daily physical activity variables; average daily steps and average daily walking and running distance (km) were created by selecting the six days of data which corresponded to the same six days aggregated in the smartphone variables. The daily activity statistics from these six days were then averaged for each measure. If a participant reported their daily walking and running distance in miles, this was converted to kilometres by multiplying the value by 1.60 before creating this average.

Lastly, BMI was calculated per person. Imperial height and weight responses were converted to metric units (centimetres and kilograms respectively). Finally, body mass index (BMI) was calculated from these values using the formula below:

Body Mass Index = Weight in kg / ((Height in cm /100) * (Height in cm /100))

5.3.3.3 *Effect size analysis*

Shapiro-Wilk normality tests were conducted on all the study variables. Only smartphone addiction scores were normally distributed. Thus, to explore if differences in smartphone measurement influenced the size of the relationships with health, Spearman correlations were conducted between all the health and smartphone variables using Fieller, Hartley & Pearsons (1957) variance when calculating 95% confidence intervals (see Table 4). Spearman correlations were also conducted between all the smartphone measures to document differences between them (see Table 6). As in with study 1, the Benjamini-Hockberg (Benjamini, & Hockberg, 1995) procedure to control for multiple comparison. As this analysis is less exploratory the rate of false discovery is changed to reelect this and become more conservative, it is changed from 10% to 5%.

Mirroring study 1, smartphone addiction scores consistently had r_S effect sizes larger than .36 with mental health, whereby estimates and objective variables were much lower (all r_S<.21) (see Fig. 3). This prompted additional analysis which assessed whether this effect size deviation across measures was statistically significant. To compare differences in the magnitude between the coefficients, we adopted Hittner, May, and Silver's (2003) modification of Dunn and Clark’s (1969) *z* test using the r package ‘cocor’ (Diedenhofen & Musch, 2015). This is suitable for the comparison of coefficient that are calculated from two dependent groups and share a variable in common (Diedenhofen & Musch, 2015). For example, it was possible using this method to compare whether the relationship between smartphone addiction and anxiety (r_S= .43) was significantly larger than the relationship between average daily screen time and anxiety (r_S = .16). We also calculated Zou (2007) confidence intervals which rejects the null hypothesis if the interval does not include 0 (Diedenhofen & Musch, 2015; Zou, 2007). Findings showed that when assessing the relationships with anxiety, depression and stress, that smartphone addiction had significantly higher correlation coefficients than any estimate and objective log variable (all *p*'s <.05) (see table 5). When examining the relationships with anxiety, depression and stress, the size of the coefficients did not differ between screen time estimates and average daily screen time (all *p*'s >05). Finally, all the correlations between estimated number of pickups and the three mental health variables were negative. These were significantly different to the three positive relationships found with mental health when assessing average daily pickups (all *p*'s <.05) (see Table 5 and Fig. 3).

Table 5.3: Descriptive statistics for study 2 variables.

Health Measures	Mean	SD	α	Smartphone Measures	Mean	SD	α
Anxiety	7.35	5.85	.94	Median Daily Screen Time (hrs)	4.62	2.30	.93
Depression	8.01	6.30	.90	Average Daily Pickups	85.76	39.94	.92
Stress	26.57	8.23	.85	Daily Screen Time Estimate (hrs)	4.38	2.15	
Body Mass Index	25.17	5.38	25.17	Daily Pickups Estimate	47.14	39.81	
Average Daily Steps	5238.07	3345.92	.84	Smartphone Addiction Scale	105.80	24.36	.92
Average Daily Walking and Running Distance	3.77	2.67	.83				

Table 5.4. Results of the Spearman correlations between smartphone and health variables from study 2.

Health Variable	Smartphone Addiction		Screen Time Estimate		Pickups Estimate		Average Daily Screen Time		Average Daily Pickups	
	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*
Mental Health										
Anxiety	.43***	0.31, 0.54	.21**	0.07, 0.35	-.08	-0.22, 0.07	.16*[n]	0.01, 0.29	.16*[n]	0.01, 0.29
Depression	.41***	0.28, 0.52	.19**	0.05, 0.32	-.10	-0.24, 0.05	.16*	0.01, 0.29	.17*	0.03, 0.31
Stress	.36***	0.23, 0.48	.21**	0.07, 0.34	-.10	-0.24, 0.04	.15*[n]	0.01, 0.29	.12	-0.02, 0.26
Physical Health										
Body Mass Index	-.07	-0.21, 0.08	.09	-0.06, 0.23	.11	-0.03, 0.25	.16*[n]	0.02, 0.30	.09	-0.5, 0.23
Average Daily Steps	-.16*	-0.30, -0.02	-.07	-0.21, 0.08	.26***	0.12, 0.39	-.07	-0.21, 0.08	.24***	0.10, 0.37
Average Daily Walking and Running Distance	-.14*	-0.28, -0.00	-.07	-0.21, 0.08	.19**	0.05, 0.33	-.09	-0.23, 0.06	.17*[n]	0.02, 0.30

Notes: * significant at $p < .05$, ** significant at $p < .01$, *** significant at $p < .0$, [n] not significant when controlling for multiple comparisons

Table 5.5. Test's comparing differences in the magnitude of the coefficients when predicting mental health from varying smartphone variables.

Variable one	Variable two	Anxiety		Depression		Stress	
		z	Zou's (2007) *CI*	z	Zou's (2007) *CI*	z	Zou's (2007) *CI*
Smartphone Addiction	Screen Time Estimate	3.14**	0.08, 0.36	3.11**	0.08, 0.36	2.10*	0.01, 0.29
Smartphone Addiction	Pickups Estimate	5.44***	0.33, 0.68	5.40***	0.33, 0.68	4.82***	0.28, 0.63
Smartphone Addiction	Average Daily Screen Time	3.48***	0.12, 0.42	3.20**	0.10, 0.40	2.65**	0.06, 0.36
Smartphone Addiction	Average Daily Pickups	3.16**	0.10, 0.43	2.80**	0.07, 0.40	2.74**	0.07, 0.41
Screen Time Estimate	Average Daily Screen Time	0.77	-0.08, 0.18	0.46	-0.10, 0.16	0.92	-0.07, 0.19
Pickups Estimate	Average Daily Pickups	-2.86**	-0.40, -0.08	-3.22**	-0.43, -0.11	-2.61**	-0.38, -0.06

Notes: * significant at $p < .05$, ** significant at $p < .01$, *** significant at $p < .001$

Table 5.6. Results of the Spearman correlations between all the smartphone variables from study 2.

Smartphone Variable	Smartphone Addiction		Screen Time Estimate		Pickups Estimate		Average Daily Screen Time		Average Daily Pickups	
	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*	Spearman r_s	Spearman 95% *CI*
Smartphone Addiction			.44***	.32, .55	.05	-.09, .19	.32***	.18, .44	.17*	.03, .31
Screen Time Estimate	.44***	.32, .55			.15*	.01, .29	.57***	.46, .66	.21**	.07, .34
Pickups Estimate	.05	-.10, .19	.15*	.01, .29			.10	-.04, .24	.30***	.16, .42
Average Daily Screen Time	.32***	.18, .44	.57***	.46, .66	.10	-.04, .24			.37***	.24, .49
Average Daily Pickups	.17*	.03, .31	.21**	.07, .34	.30***	.16, .42	.37***	.24, .49		

Notes: * significant at $p < .05$, ** significant at $p < .01$, *** significant at $p < .001$

Table 5.7. Results of linear regression models with health measures as the criterion variables, and smartphone measures as predictors.

	B. with criterion variable						*B. with criterion variable (only objective measures)*					
Model	Anx	Dep	Stress	BMI	Steps	Dist	Anx	Dep	Stress	BMI	Steps	Dist
Intercept	-2.95	-3.65	14.06***	25.05***	6061.84***	4.47***	5.46***	5.55***	24.15***	23.40***	4009.15***	3.35***
Average Daily Screen Time	-0.10	0.04	0.07	0.20	-36.04	-0.13	0.25	0.34	0.45	0.33	-129.31	-0.14
Average Daily Pickups	0.01	0.00	0.00	-0.00	20.20**	0.01*n	0.01	0.01	0.00	0.00	21.29***	0.01*
Screen Time Estimate	0.14	-0.03	0.05	0.45*n	-44.14	0.07						
Pickups Estimate	-0.01	-0.00	-0.01	0.02	10.01	0.00						
Smartphone Addiction	0.10***	0.11***	0.12***	-0.03	-25.22*n	-0.02						

R^2	.18	.17	.13	.06	.10	.05	.02	.02	.02	.02	.06	.03
R^2_{Adj}	.16	.15	.11	.04	.08	.03	.01	.02	.01	.01	.05	.02

Notes: R^2_{Adj} = Adjusted R^2, B = beta estimates, * beta estimates significant to $p < .05$, ** beta estimates significant to $p < .01$, *** beta estimates significant to $p < .001$. Anx = Anxiety, Dep = Depression, Stress = Stress, BMI = Body Mass Index, Steps = Average Daily Steps, Dist = Average Daily Walking and Running Distance, All VIF scores between 1 – 2 and all tolerance scores > .59, ^{n} = Not significant when controlling for multiple comparisons.

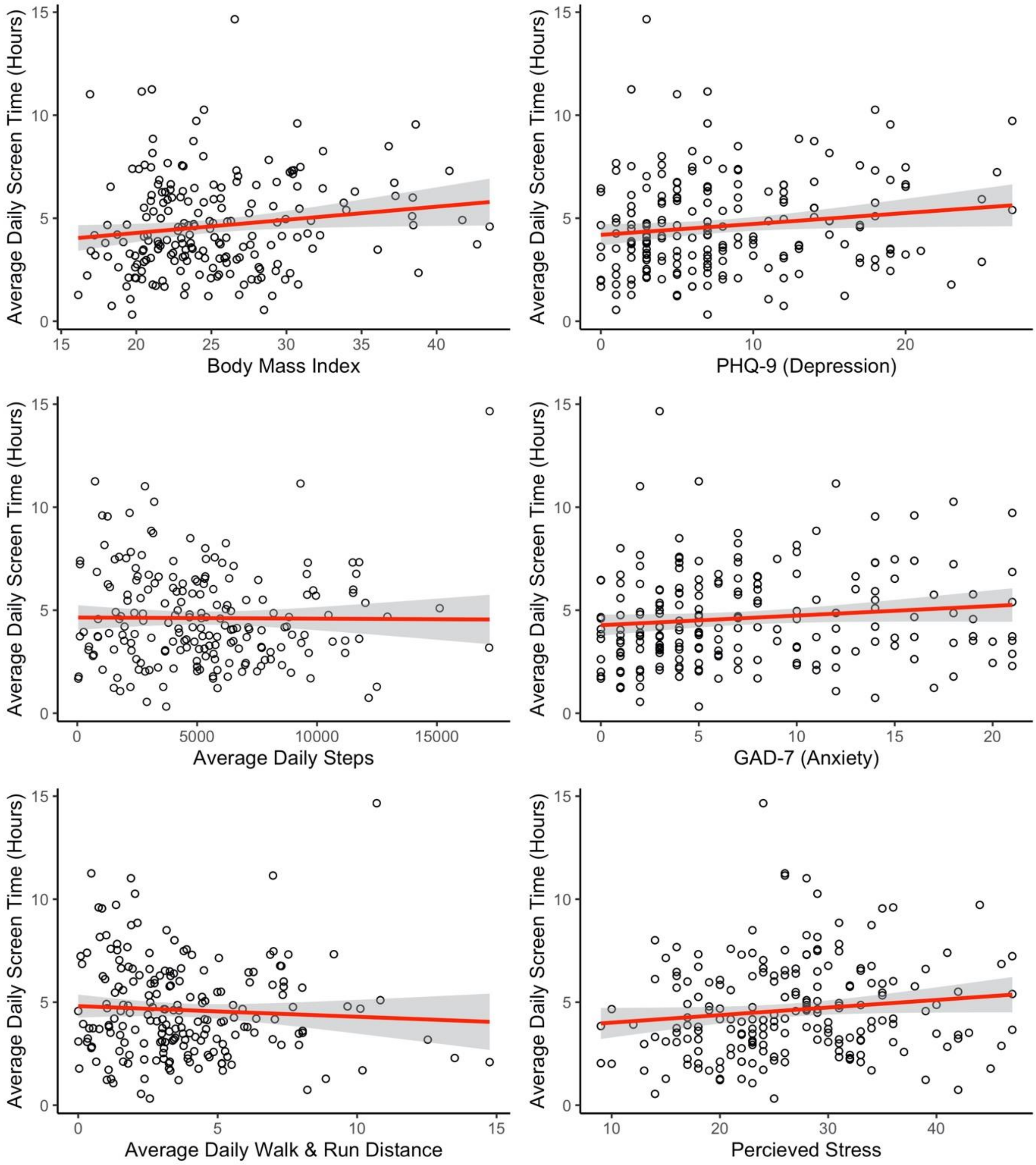


Figure 5.2. Scatter plots showing the linear relationships between average daily screen time (hrs) with six health variables; body mass index, averaged daily steps, average daily walking and running distance, anxiety, depression and stress. The red regression line represents the linear relationship between the two, and the surrounding grey area represents the 95% confidence interval.

Exploratory analysis – Tests of difference between groups with low and high mental health symptomology.

Measuring 'percentage variance explained' through the exploration of effect sizes has been criticised in recent work, advocating that significance testing between groups is a better indicator of whether screen time impacts mental health (Twenge, 2019). While this approach is in contradiction to many other statistical recommendations (Cumming, 2014), it was of interest to explore whether our conclusions would differ if we adopted this type of analysis. Consequently, as the GAD-7 and PHQ-9, have 'cut off points' which indicate if people are at risk of having each disorder, we used these to create two groups; a 'low risk' or a 'high risk' group. Meaning a GAD-7 score of 9 or higher and a PHQ-9 score of 10 or higher. These measures have high sensitivity and specificity (both > .80) when diagnosing depression and anxiety disorders (Kroenke et al., 2001; 2007). However, due to lack of further psychological assessment those who exceeded the defined cut-off points for each disorder were considered at risk, rather than defined as having the disorder. We then examined if people experienced different levels of daily smartphone use and PSU dependent on what group they belonged to.

To create groups for the analysis, the 50 people who were considered 'high risk' for both anxiety and depression were collated into one group. This group used their phone for an average of 4.72 hours a day (SD = 2.27) and picked up their phone on average 84.20 times a day (SD = 37.98). Those who didn't exceed the cut-off values for either condition (scored less than 10 on both scales) were placed in a 'low risk' group (n = 124). This group used their phone for an average of 4.41 hours a day (SD = 2.25) and picked up their phone on average 84.07 times a day (SD = 42.55). Wilcoxon rank sum tests showed that the two groups did not significantly differ in their amounts of average daily screen time [*W* = 3357, *p* = .39] or average daily pickups [*W* = 3216, *p* = .70]. This was mirrored when exploring differences in estimated daily screen time [*W* = 3489.5, *p* = .19] and estimated daily pickups [*W* = 2721, *p* = .20]. Therefore, those who were 'high risk' of having both general anxiety disorder and major depression did not use their smartphone's differently to those who were 'low risk' for both conditions. However, a significant difference was found between the two groups on levels of smartphone addiction [*W* = 4505.5, *p* < .001], this was reliable even when controlling for multiple comparisons. Specifically, the 'at risk' group had higher smartphone addiction scores [*M* = 116, *SD* = 23.67] than the 'low risk' group [*M* = 98.91, *SD* = 21.91]. Consequently, if smartphone use is measured with subjective estimates or objective logs, we find no difference between 'high risk' and 'low risk' groups in terms of usage. However, if confounding usage and PSU, one would conclude the opposite if measuring 'usage', via a smartphone addiction scale, positing that those with mental health symptomology have higher usage.

5.3.3.4 ***Exploratory analysis – Linear Regression Models***

Many researcher's build regression models to investigate of there is a linear or logarithmic relationship between health and smartphone measures (Csibi, et al., 2018; David, Roberts, & Christenson, 2018; Kim et al., 2016; Regan et al., 2020; Richardson, Hussain, & Griffiths, 2018). Following suit, it was of interest to see whether models predicting mental health symptomology would be influenced by the type of smartphone variables included as predictors. Notably, when including all five smartphone measures in models, only smartphone addiction scores significantly predicted mental health scores (see Table 7). Furthermore, models which only contained objective smartphone measures showed no significant relationships with mental health (all $R^2 \leq .02$, all *p*'s >.05). Finally, average daily pickups significantly predicted average daily steps and average daily walking and running distance across models (see Table 7).

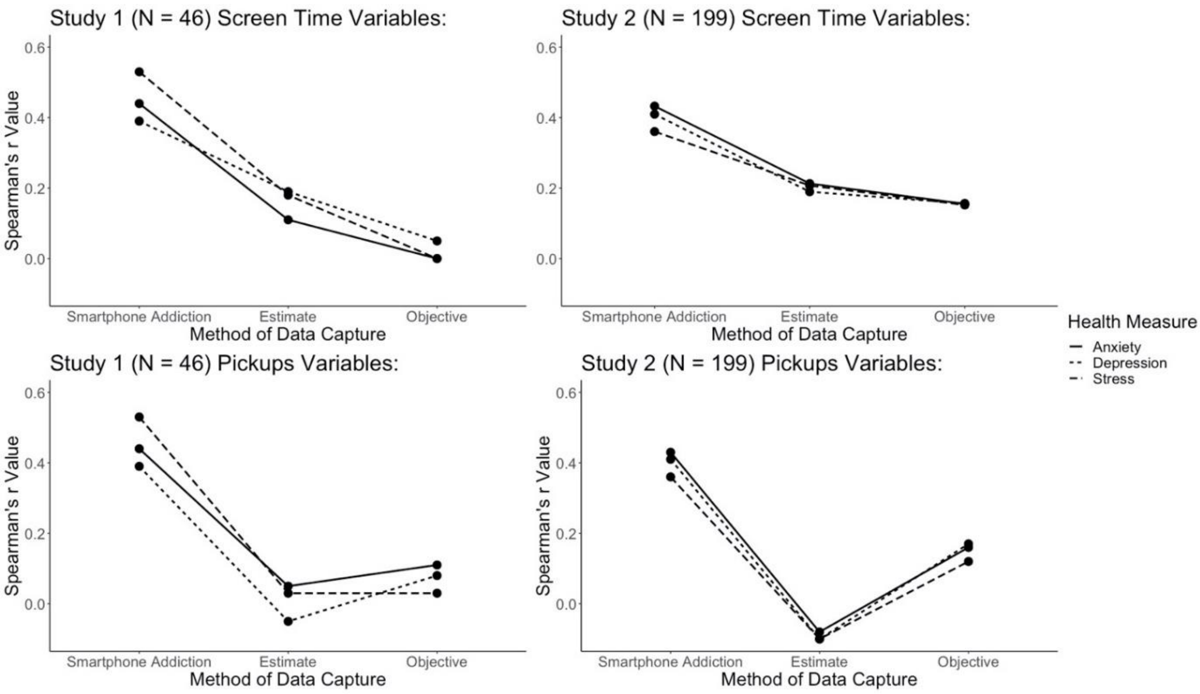


Figure 5.3: Graphs showing how a change in measurement effects the relationships between smartphone use and depression, anxiety and stress across both studies. The top row represents how smartphone addiction scores compare to estimated and actual screen time when correlating with mental health. The bottom row represents how smartphone addiction scores compare to estimated and actual pickups when correlating with mental health.

5.4 General Discussion

Here, we considered whether different ways of measuring 'smartphone use', notably through problematic smartphone usage (PSU) scales, subjective estimates, and objective logs, leads to contrasting associations with health. Across two samples including iPhone (n=199) and Android (n=46) users, we observed that measuring smartphone interactions with PSU scales produced larger effect sizes with mental health than subjective estimates or objective logs. Notably, the size of the relationship was fourfold in study 1, and over twice as large in study 2 when employing a smartphone addiction scale in comparison to objective measures. Moreover, r_S < .17 was repeatedly found between objective smartphone use (daily pickups and screen time) and mental health symptomology (anxiety, depression, and stress), whereas bigger effect sizes were found with problematic usage scale (all r_S >.36). Hence, decisions concerning definitions and subsequent measurements of 'use' can hugely influence the size of the relationships with mental wellbeing.

Thus, it can be argued that existing conclusions linking mental health to general smartphone use are likely premature and overstated, given a recent review showed 70% of studies adopt PSU scales (Thomée 2018). Notably, this review concluded that intense or frequent mobile phone use was associated with greater mental health symptomology, yet this conclusion was based primarily on findings from PSU studies. It is this confounding between a person's appraisal of their smartphone use and their actual usage which is troubling given they have vastly different relationships with mental wellbeing. To address this, we explored direct measures of usage and showed that the size of the relationships between objective measures and mental health is on average r_S= 0.15 (all <.17). To put into context, this is lower than the average effect size found across many psychology studies (r = .21), just less than the relationship between Nicotine patch (vs. placebo) and smoking abstinence (r = .18), and about the same size as the relationship between post-high school grades and job performance (r = .16) (Funder & Ozer, 2019; Meyer et al., 2001). When adjusting for new recommendations that 'small', 'typical', and 'relatively large' effects fall around *r* coefficients of ~.10, ~.20 and ~.30, respectively (Gignac & Szodorai, 2016), it has been suggested that if social media had *"destroyed our lives"*, then moderate to large effects would be found (*r* >.20) (Appel, Marker, & Gnambs, 2020, pp.62). Using this benchmark our findings suggest that general smartphone use does not have extreme or profound effects on wellbeing, against recent claims (Twenge, 2017).

This was further supported in our regression analysis which showed that average daily pickups and average daily screen time did not significantly predict anxiety, depression or stress, and explained ~2% of the variance. Additionally, those who exceeded the 'cut off point' for either general anxiety disorder or major depressive did not use their phone significantly more than those who scored below the threshold. Contemporary research is also showing similar findings. For instance, in a large sample of New Zealand adults (*n* = 19,075), the association between social media use and wellbeing was found to be very weak (Stronge et al., 2019). In another study, 'intense' general smartphone use did not predict negative wellbeing when using objective logs (Katevas et al., 2018). When using specification curve analysis to examine self-reports from a large sample of adolescents (*n* = 355,358), the association between digital technology use and wellbeing was found to be small, explaining .4% of the variance (Orben & Przybylski, 2019). As we have also found that increased objective screen time and pickups explained 2% or less of the variance in mental health measures, it questions whether reducing smartphone use should be a priority for public health interventions.

Existing research also coincides with our pattern of results which show that PSU scales have stronger relationships with depression, anxiety and stress than estimates and objective logs of use (Rozgonjuk et al. 2018; Vahedi & Saiphoo, 2018). In a similar study which measured objective smartphone screen time over a weeklong period, findings showed that average daily depressive mood positively correlated with smartphone addiction scores, yet objective screen time minutes were not related to depression and anxiety (Rozgonjuk et al. 2018). It can be speculated that scores from PSU scales have stronger relationships with mental health for several reasons. Firstly, in line with the cognitive behavioural approach (Beck, 1967), one could argue that people's negative appraisals of their smartphone use could influence their mood states, and that problematic usage scales capture these thoughts and beliefs. Mirroring this, a recent study showed that higher levels of the trait mindfulness (being aware in the present moment) was associated with lower rates of smartphone addiction and nomophobia scores but was not associated with objective smartphone logs (Regan et al., 2020). When conducting regression analysis, findings from study 2 showed that smartphone addiction was the only smartphone variable which significantly predicted anxiety, depression and stress when including all five 'usage' measures in the same model. This suggests that reducing people's worries towards their smartphone use is likely to have greater mental health benefits than reducing the use of the device itself.

However, some have argued that a very large effects of r = .40 in psychology studies are likely to be likely to be an overestimate, and warrant some scepticism (Funder & Ozer, 2019). Following suit, a quick analysis of the data from study 2 showed that the relationship between anxiety and smartphone addiction was the same as the relationship between height and weight (both r_s= .43). This large relationship may be due to the overlap between PSU and mental health constructs, illustrated in recent factor analysis research (Davidson, Shaw, & Ellis, 2020). Notably, individual questions in the SAS which enquired about emotive states appeared in factors with stress and depression items (Davidson, Shaw, & Ellis, 2020). Hence, it could be argued that these cross-loadings between PSU and mental health could artificially enhance relationships due to a lack of independence. It is further possible that 'method bias' influenced the size of the correlation coefficients because the scales are similar to each other in terms of how the items are worded (Podsakoff, MacKenzie, & Podsakoff, 2012). Notably, every question in the SAS assesses a percieved problem, echoing mental health scales (Kwon et al. 2013; Spitzer, et al., 2006; Kroenke, et al., 2001). However, this negative wording in itself could be a further source of bias. For example, it has been shown that correlations between role conflict, role ambiguity and other constructs reduced by 238% when controlling for wording effects, by balancing the number of positively and negatively worded questions. Thus, in line with recent criticims, issues with how PSU scales are developed and conceptualised could be inflating the relationships with mental health in comparison to objective usage measures (Davidson, Shaw, & Ellis, 2020).

Furthermore, while some have found participants with higher smartphone addiction scores have lower muscle mass (Kim, Kim, & Jee, 2015), our findings derived from objective logs are less conclusive. Neither study found a relationship between body mass index and objective smartphone use is incoherent across studies when controlling for multiple comparisons. There initially appeared to be a positive relationship between average daily steps and average daily walking and running distance with objective daily pickups but when controlling for multiple comparisons this was deemed to be a spurious relationship. The portable nature of the smartphone may has been suggested as an advantageous to the health of the user. Arora et al., (2013) found that computer use, tv viewing and video gaming were associated with increased BMI, but conversely, did not find the same for mobile phone use. They stated, *"the portable nature of a mobile telephone does not require the user to remain in one place during use, thus allowing movement"* (Arora, et al. 2013, pp. 1258). This is in line with recent discussions that screen time is often conceptualised in absence of 'exergaming' and other activities which involve physical activity whilst engaging with the device (Kaye, et al., 2020). Future research should therefore adopt a more nuanced approach and understand both the costs and benefits of specific applications which can be monitored via newly developed apps (Geyer, et al., 2020). Recent work has shown that total time spent using smartphones had r = .16 effect sizes with anxiety and depression, but that certain categories of apps had beneficial relationships (e.g. time spent reading books) (David et al., 2018). Therefore, claiming all smartphone use as negative oversimplifies a very complex and multifaceted phenomenon.

However, it can become difficult to objectively measure the use of a specific application across many devices (e.g. documenting time spent on Netflix across smartphones, televisions and tablets) (Kaye et al., 2020). In these cases, researchers may document estimates of use instead. Though importantly, our findings suggest there is a discrepancy between usage statistics documented from objective logs vs subjective estimates (see Table 6), which has been found repeatedly across many studies (Andrews et al., 2015; Boase & Ling, 2013; Parslow, Hepworth, & McKinney, 2003; Ellis et al., 2019; Kobayashi & Boase, 2012; Lee, et al., 2017; Vrijheid et al., 2006).. Across study 2 and previous work, estimated frequency of 'pickups' had greater deviation from its objective counterpart than daily estimates of screen time (Andrews, et al., 2015; Ellis, et al., 2019). In accordance, relationships between smartphone use and health significantly differed dependent on whether estimated pickups or objective pickups were measured. Thus, if subjective estimates are to be collected, it is advocated that researchers start factoring this measurement error into statistical models, as we have quantified here the size of this error, e.g. through regression calibration, multiple imputation or latent class analysis (Ellis, 2019; van Smeden, Lash and Groenwold, 2019). That way, results are likely to be closer to what is found with objective measures and could act as a better proxy.

5.4.1 *Limitations*

Both studies are cross-sectional; therefore, we cannot make any causal claims regarding the impact of smartphone use and mental health. However, by using a quasi-experimental approach in the exploratory analysis of study 2 and through analysing the naturally occurring levels of mental health symptomology in our sample, our findings cast doubt on the presence of causal relationships, as those in a high symptomology group did not have increased general smartphone usage. It is further possible that participants may have received feedback from Apple Screen Time prior to the study, which would have influenced their estimation of use. The size of the relationship between estimated screen time and actual screen time is larger in study 2 than previous work and may explain why effect sizes with mental health did not significantly differ between these measures (Andrews, et al., 2015; Ellis, et al., 2019). However, as this correlation is below an acceptable threshold of r_S=.8, there is still an element of error between actual and self-reported screen time, which requires controlling for in future analysis.

In addition, by moving the study to an online platform, we achieved a larger and more representative sample. However, this meant losing some of the precision obtained using laboratory based bioimpedance measures when examining physical health. Nonetheless, as BMI scores in study 2 had large correlations with body fat percentage (r_S = .70) and skeletal muscle mass % (r_S = -.73) we accepted this as a relatively good proxy. Furthermore, as self-reports of height and weight may also have measurement error, we analysed the ranges of BMI values that were reported in study 2. Findings showed that our sample had BMI values that were in line with what is expected in representative sample (WHO, 2018). However, future research would benefit from exploring ways to gather body composition such as body fat percentage in online settings.

5.5 Conclusions

To conclude, choosing between measurement tools, and accepting the benefits and limitations of that choice is an unavoidable facet of all research. However, when understanding or making claims regarding the effects of a particular behaviour on health, the cost of this error can be considerable. Here we demonstrate that problematic smartphone usage scales have significantly larger relationships with mental health than objective logs of use. These are over twice the size in a large sample and over four-fold in a small sample. Thus, objectivity should not only be present in the way we measure health but should also reside in the way we measure usage (e.g. in neurological studies) (Hutton, et al., 2019). Without this, we risk perpetuating misleading conclusions into public understanding, which in itself may cause harm. Specifically, the concept of 'problematic use' requires stringent examination, given that increased objective screen time and pickups does not relate to mental health symptomology in regression analysis, and due to issues with conceptualisation and scale development. This is in agreement with recent arguments stating that 'excessive 'smartphone usage does not neccesaily equate to 'problematic use '(Billieux, Philippot, et al., 2015; Panova & Carbonell, 2018). Consequently, PSU scales may only capture people's appraisals of their smartphone use, rather than an underlying pathology. Our findings would actually favour addressing people's concerns about their usage over reducing their overall device use, as this relates more strongly mental health symptomology. Thus, even if people do have specific worries in relation to their smartphone use, our data implies that limiting all smartphone use to only a certain amount per day is unlikely to have any demonstrable benefits and should not be a priority for public health interventions at this time.

References

Andrews, S., Ellis, D. A., Shaw, H., & Piwek, L. (2015). Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use. *Plos One*, *10*, e0139004. https://doi.org/10.1371/journal.pone.0139004

Appel, M., Marker, C., & Gnambs, T. (2020). Are Social Media Ruining Our Lives? A Review of Meta-Analytic Evidence. *Review of General Psychology*, *24*(1), 60–74. https://doi.org/10.1177/1089268019880891

Arora, T., Hussain, S., Hubert Lam, K. B., Lily Yao, G., Neil Thomas, G., & Taheri, S. (2013). Exploring the complex pathways among specific types of technology, self-reported sleep duration and body mass index in UK adolescents. *International Journal of Obesity*, *37*(9), 1254–1260. https://doi.org/10.1038/ijo.2012.209

Barkley, J. E., & Lepp, A. (2016). Mobile phone use among college students is a sedentary leisure behavior which may interfere with exercise. *Computers in Human Behavior*, *56*, 29–33. https://doi.org/10.1016/j.chb.2015.11.001

Beck, A. T. (1967). *Depression: Causes and Treatment*. Philadelphia: University of Pennsylvania Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.

Billieux, J., Maurage, P., Lopez-Fernandez, O., Kuss, D. J., & Griffiths, M. D. (2015). Can Disordered Mobile Phone Use Be Considered a Behavioral Addiction? An Update on Current Evidence and a Comprehensive Model for Future Research. *Current Addiction Reports*, *2*, 156–162. https://doi.org/10.1007/s40429-015-0054-y

Billieux, J., Philippot, P., Schmid, C., Maurage, P., De Mol, J., & Van der Linden, M. (2015). Is Dysfunctional Use of the Mobile Phone a Behavioural Addiction? Confronting Symptom-Based Versus Process-Based Approaches. *Clinical Psychology and Psychotherapy*, *22*(5), 460–468. https://doi.org/10.1002/cpp.1910

Bishara, A. J., & Hittner, J. B. (2017). Confidence intervals for correlations when data are not normal. *Behavior Research Methods*, *49*(1), 294–309. https://doi.org/10.3758/s13428-016-0702-8

Boase, J., & Ling, R. (2013). Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*, *18*(4), 508–519. https://doi.org/10.1111/jcc4.12021

British Psychological Society (2018). *Code of Ethics and Conduct (2018)*. Retrieved May 09, 2020 from: https://www.bps.org.uk/news-and-policy/bps-code-ethics-and-conduct

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, *26*(1), 17–31. https://doi.org/10.1016/j.cpr.2005.07.003

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Percieved Stress. *Journal of Health and Social Behavior*, *24*(4), 385–396.

Csibi, S., Griffiths, M. D., Cook, B., Demetrovics, Z., & Szabo, A. (2018). The Psychometric Properties of the Smartphone Application-Based Addiction Scale (SABAS). *International Journal of Mental Health and Addiction*, *16*(2), 393–403. https://doi.org/10.1007/s11469-017-

9787-2

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2019). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in Human Behavior*, *104,* 106160. https://doi.org/10.1016/j.chb.2019.106160

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

David, M. E., Roberts, J. A., & Christenson, B. (2018). Too Much of a Good Thing: Investigating the Association between Actual Smartphone Use and Individual Well-Being. *International Journal of Human-Computer Interaction,* *34*(3), 265–275. https://doi.org/10.1080/10447318.2017.1349250

Davidson, B. I., Shaw, H., & Ellis, D. A. (2020). Fuzzy Constructs in Assessment: The Overlap between Mental Health and Technology "Use". *PsyArXiv*. https://doi.org/10.31234/osf.io/6durk

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE,* *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Elhai, J. D., & Contractor, A. A. (2018). Examining latent classes of smartphone users: Relations with psychopathology and problematic smartphone use. *Computers in Human Behavior*, *82*, 159–166. https://doi.org/10.1016/j.chb.2018.01.010

Elhai, J. D., Dvorak, R. D., Levine, J. C., & Hall, B. J. (2017). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders, 207,* 251–259. https://doi.org/10.1016/j.jad.2016.08.030

Elhai, J. D., Yang, H., Rozgonjuk, D., & Montag, C. (2020). Using machine learning to model problematic smartphone use severity: The significant role of fear of missing out. *Addictive Behaviors*, *103*, 106261. https://doi.org/10.1016/j.addbeh.2019.106261

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*, *97*, 60–66. https://doi.org/10.1016/j.chb.2019.03.006

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior? *International Journal of Human-Computer Studies*, *130*, 86–92. https://doi.org/10.1016/j.ijhcs.2019.05.004

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Fieller, E. C., Hartley, H. O., & Pearsons, E. S. (1957). Tests for rank correlation coefficients: I. *Biometrika*, *44*, 470–481.

Genc, Z. (2014). Parents ’Perceptions about the Mobile Technology Use of Preschool Aged Children. *Procedia - Social and Behavioral Sciences*, *146*, 55–60. https://doi.org/10.1016/j.sbspro.2014.08.086

Geyer, K. (2018). *activity logger* [Software]. Retrieved from https://play.google.com/store/apps/details?id=sensor.acitivty.android.activitylog&hl=en_GB

Geyer, K., Ellis, D. A., Shaw, H., & Davidson, B. I. (2020). Open source smartphone apps and tools for measuring, quantifying and visualizing screen time. *PsyArXiv*. https://doi.org/10.31234/osf.io/eqhfa

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Harwood, J., Dooley, J. J., Scott, A. J., & Joiner, R. (2014). Computers in Human Behavior Constantly connected – The effects of smart-devices on mental health. *Computers in Human Behavior*, *34*, 267–272. https://doi.org/10.1016/j.chb.2014.02.006

Hamilton, M.T., Hamilton, D.G., & Zderic, T.W. (2007). Role of Low Energy Expenditure and Sitting in Obesity, Metabolic. *Diabetes*, *56*(11), 2655–2667. https://doi.org/10.2337/db07-0882.CVD

Heffer, T., Good, M., Daly, O., MacDonell, E., & Willoughby, T. (2019). The Longitudinal Association Between Social-Media Use and Depressive Symptoms Among Adolescents and Young Adults: An Empirical Reply to Twenge et al. (2018). *Clinical Psychological Science*, 1–9. https://doi.org/10.1177/2167702618812727

Hewson, C., Buchanan, T., Brown, I., Coulson, N., Hagger-Johnson, G., Joinson, A., … Oates, J. (2013). Ethics Guidelines for Internet-mediated Research. In *INF206/1.2013*. Retrieved from www.bps.org.uk/publications/policy-and-guidelines/research-guidelines-policy-documents/research-guidelines-poli

Hittner, J. B., May, K., & Silver, N. C. (2003). A monte carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

Hutton, J. S., Dudley, J., Horowitz-Kraus, T., DeWitt, T., & Holland, S. K. (2019). Associations Between Screen-Based Media Use and Brain White Matter Integrity in Preschool-Aged Children. *JAMA Pediatrics*, e193869–e193869.

https://doi.org/10.1001/jamapediatrics.2019.3869

Johannes, N., Meier, A., Reinecke, L., Ehlert, S., Setiawan, D. N., Walasek, N., … Veling, H. (2019). The Relationship Between Online Vigilance and Affective Well-Being in Everyday Life: Combining Smartphone Logging with Experience Sampling. *Preprint.* https://doi.org/10.31234/osf.io/t3wc2

Katevas, K., Arapakis, I., & Pielot, M. (2018). Typical Phone Use Habits: Intense Use Does Not Predict Negative Well-Being. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services.* https://doi.org/10.1145/3229434.3229441

Kaye, L. K., Orben, A., Ellis, D. A., Hunter, S. C., & Houghton, S. (2020). The conceptual and methodological mayhem of "screen time." *International Journal of Environmental Research and Public Health*, *17*(10). https://doi.org/10.3390/ijerph17103661

Kim, J.-H. (2017). Smartphone-mediated communication vs. face-to-face interaction: Two routes to social support and problematic use of smartphone. *Computers in Human Behavior*, *67*, 282–291. https://doi.org/10.1016/j.chb.2016.11.004

Kim, J., Hwang, Y., Kang, S., Kim, M., Kim, T.-S., Kim, J., … Park, S. (2016). Association between Exposure to Smartphones and Ocular Health in Adolescents. *Ophthalmic Epidemiology*, *23*(4), 269–276. https://doi.org/10.3109/09286586.2015.1136652

Kim, S.-E., Kim, J.-W., & Jee, Y.-S. (2015). Relationship between smartphone addiction and physical activity in Chinese international students in Korea. *Journal of Behavioral Addictions*, *4*(3), 200–205. https://doi.org/10.1556/2006.4.2015.028

Kobayashi, T., & Boase, J. (2012). No Such Effect? The Implications of Measurement Error in Self-Report Measures of Mobile Communication Use. *Communication Methods and*

Measures, *6*(2), 126–143. https://doi.org/10.1080/19312458.2012.679243

Kroenke, K, Spitzer, R. L., Janet, B. W., & Williams, D. S. W. (2001). The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, Kurt, Spitzer, R. L., Janet, B. W., Williams, D. S. ., Monahan, P., & Löwe, B. (2007). Anxiety Disorders in Primary Care: Prevalence, Impairment, Comorbidity, and Detection. *Annals of Internal Medicine*, *146*, 317–325.

Kwon, M., Lee, J. Y., Won, W. Y., Park, J. W., Min, J. A., Hahn, C., … Kim, D. J. (2013). Development and Validation of a Smartphone Addiction Scale (SAS). *PLoS ONE*, *8*(2). https://doi.org/10.1371/journal.pone.0056936

Lee, H., Ahn, H., Nguyen, T. G., Choi, S. W., & Kim, D. J. (2017). Comparing the self-report and measured smartphone usage of college students: A pilot study. *Psychiatry Investigation*, *14*(2), 198–204. https://doi.org/10.4306/pi.2017.14.2.198

Lepp, A., Barkley, J. E., Sanders, G. J., Rebold, M., & Gates, P. (2013). The relationship between cell phone use, physical and sedentary activity, and cardiorespiratory fitness in a sample of U.S. college students. *International Journal of Behavioral Nutrition and Physical Activity*, *10*, 1–9. https://doi.org/10.1186/1479-5868-10-79

Loid, K., Täht, K., & Rozgonjuk, D. (2020). Do pop-up notifications regarding smartphone use decrease screen time, phone checking behavior, and self-reported problematic smartphone use? Evidence from a two-month experimental study. *Computers in Human Behavior*, *102*(August 2019), 22–30. https://doi.org/10.1016/j.chb.2019.08.007

McDonald, J. H. (2014). Handbook of Biological Statistics (3rd ed.). Sparky House Publishing. Baltimore, Maryland.

Meyer, G. K., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. I., Dies, R. R., … Reed, G. M. (2001). Psychological Testing and Psychological Assessment: A Review of Evidence and Issues. *American Psychologist Copyright*, *56*(2), 128–165. https://doi.org/10.1037//OOO3-O66X.56.2.128

Orben, A. (2018). Cyberpsychology: A Field Lacking Theoretical Foundations. *PsyPAG The Quarterly*, (105).

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*, 173–182. https://doi.org/10.1038/s41562-018-0506-1

Panova, T., & Carbonell, X. (2018). Is smartphone addiction really an addiction? *Journal of Behavioral Addictions*, *7*(2), 252–259. https://doi.org/10.1556/2006.7.2018.49

Parslow, R. C., Hepworth, S. J., & McKinney, P. A. (2003). Recall of past use of mobile phone handsets. *Radiation Protection Dosimetry*, *106*(3), 233–240. https://doi.org/10.1093/oxfordjournals.rpd.a006354

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Regan, T., Harris, B., Van Loon, M., Nanavaty, N., Schueler, J., Engler, S., & Fields, S. A. (2020). Does mindfulness reduce the effects of risk factors for problematic smartphone use? Comparing frequency of use versus self-reported addiction. *Addictive Behaviors*, *108*, 106435. https://doi.org/10.1016/j.addbeh.2020.106435

Richardson, M., Hussain, Z., & Griffiths, M. D. (2018). Problematic smartphone use, nature connectedness, and anxiety. *Journal of Behavioral Addictions*, *7*(1), 109–116.

https://doi.org/10.1556/2006.7.2018.10

Rozgonjuk, D., Levine, J. C., Hall, B. J., & Elhai, J. D. (2018). The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. *Computers in Human Behavior*, *87*, 10–17. https://doi.org/10.1016/j.chb.2018.05.019

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. Journal of medical Internet research, 17(7), e4273.

Science-and-Technology-Commitee. (2019). *Impact of social media and screen-use on young people's health.* Retrieved from https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf

Shaw, H., Ellis, D. A., & Ziegler, F. V. (2018). The Technology Integration Model (TIM). Predicting the continued use of technology. *Computers in Human Behavior*, *83*, 204-214. https://doi.org/10.1016/j.chb.2018.02.001

Spitzer, R. ., Kroenke, K., Janet, B. W., Williams, D., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stich, J. F., Tarafdar, M., Stacey, P., & Cooper, S. C. (2019). Appraisal of email use as a source of workplace stress: A person-environment fit approach. *Journal of the Association for Information Systems*, *20*(2), 132–160. https://doi.org/10.17705/1jais.00531

Stronge, S., Mok, T., Ejova, A., Lee, C., Zubielevitch, E., Yogeeswaran, K., … Sibley, C. G. (2019). Social Media Use Is (Weakly) Related to Psychological Distress. *Cyberpsychology, Behavior, and Social Networking*, *22*(9), 604–609. https://doi.org/10.1089/cyber.2019.0176

Thomée, S. (2018). Mobile Phone Use and Mental Health. A Review of the Research That Takes a Psychological Perspective on Exposure. *International Journal of Environmental Research and Public Health*, *15*(12), 2692. https://doi.org/10.3390/ijerph15122692

Twenge, J. M. (2017). Have Smartphones destroyed a generation? Retrieved October 14, 2019, from https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/

Twenge, J. M., Martin, G. N., & Campbell, W. K. (2018). Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology. *Emotion*, *18*(6), 765–780. https://doi.org/10.1037/emo0000403

Twenge, J. M. (2019). More Time on Technology, Less Happiness? Associations Between Digital-Media Use and Psychological Well-Being. *Current Directions in Psychological Science*, *28*(4), 372–379. https://doi.org/10.1177/0963721419838244

van Smeden, M., Lash, T.L., & Groenwold, R .H. (2019). Reflection on modern methods: five myths about measurement error in epidemiological research. *International Journal of Epidemiology*, dyz251. https://doi.org/10.1093/ije/dyz251

Vahedi, Z., & Saiphoo, A. (2018). The association between smartphone use, stress, and anxiety: A meta-analytic review. *Stress and Health*, *34*(3), 347–358. https://doi.org/10.1002/smi.2805

Verney, J., Schwartz, C., Amiche, S., Pereira, B., & Thivel, D. (2015). Comparisons of a Multi-Frequency Bioelectrical Impedance Analysis to the Dual-Energy X-Ray Absorptiometry Scan in Healthy Young Adults Depending on their Physical Activity Level. *Journal of Human Kinetics*, *47*(1), 73–80. https://doi.org/10.1515/hukin-2015-0063

Vrijheid, M., Cardis, E., Armstrong, B. K., Auvinen, A., Berg, G., Blaasaas, K. G., …

Woodward, A. (2006). Validation of short term recall of mobile phone use for the Interphone study. *Occupational and Environmental Medicine*, *63*(4), 237–243. https://doi.org/10.1136/oem.2004.019281

WHO. (2018). Obesity and overweight. Retrieved September 23, 2019, from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Xiang, M. Q., Lin, L., Wang, Z. R., Li, J., Xu, Z., & Hu, M. (2020). Sedentary Behavior and Problematic Smartphone Use in Chinese Adolescents: The Moderating Role of Self-Control. *Frontiers in Psychology*, *10*, 1–9. https://doi.org/10.3389/fpsyg.2019.03032

Yang, Z., Asbury, K., & Griffiths, M. D. (2019). "A Cancer in the Minds of Youth?" A Qualitative Study of Problematic Smartphone Use among Undergraduate Students. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-019-00204-z

Zagalaz-Sánchez, M. L., Cachón-Zagalaz, J., Sánchez-Zafra, M., & Lara-Sánchez, A. (2019). Mini review of the use of the mobile phone and its repercussion in the deficit of physical activity. *Frontiers in Psychology*, *10*, 1307. https://doi.org/10.3389/fpsyg.2019.01307

Zou, G. Y. (2007). Toward Using Confidence Intervals to Compare Correlations. *Psychological Methods*, *12*(4), 399–413. https://doi.org/10.1037/1082-989X.12.4.399

Chapter VII

Conclusion

This thesis provides evidence that, to date, psychology has perhaps only scratched the surface when it comes to using smartphones within mainstream research. Smartphones can capture a variety of data for psychology including: the context of behaviour (C*hapter II*; PEG

LOG) and digital behaviour (C*hapter III*; Usage Logger). Such software can be employed to deliver strong contributions for applied psychology. For instance, the thesis identified that past estimates of smartphone usage poorly correlate with actual usage (C*hapter IV*) and used that tool to falsify past research demonstrating links between smartphone usage and negative mental and physical health outcomes (*Chapter V*). The development of smartphone apps and the utilisation of them in research has led me to disagree with early theories about how smartphone apps will influence psychological research.

This concluding chapter will briefly outline the potential theoretical contributions that smartphones can have then engage primarily in an evaluation the contributions of psych apps to psychological research by incorporating evidence provided in the thesis alongside other research. Psych apps may be a significant improvement upon existing methodologies, however there are extensive issues for the reliability of validity of data for studies conducted now and in the future. If these obstacles can be overcome then psych apps and smartphones represent an objective data source of actual behaviour which occurred ecologically. With such a data source psychological science will have inherited a solid basis in evidence from this psychology should generate more accurate theories and have far greater potential to meaningfully impact society. Therefore, I will argue for embracing server-less apps should aim to collect data reliably, that is validated, and yet remain transparent in their operation.

7.1 Theoretical importance of Psych apps

This section intends to provide support for the assertion that psych apps will deliver theoretical freedom to psychologists as the increase in computational power has done by the invention of the computer has done achieved (see *Chapter I*). Generally, psychologists were limited by their ability to conduct statistical tests by how long it took to calculate them. Non-parametric tests involved an extensive amount of work to calculate manually and thus there was an incentive to assume psychological phenomena was normal and therefore parametric tests could be used to analyses data.

Multiple methodologies in psychology require researchers to adopt assumptions, otherwise the underlying logic of the methodology is compromised. For instance, psychological tools such as the highly popular NEO inventory (McCrae & Costa, 2008) assume that individuals are rational or have bounded rationality (Simon, 1957a; Simon 1957b; Costa & McCrea, 1999). Here they define rationality in terms of self-reflection "Rationality is the assumption that people are generally capable of understanding themselves and other (cf. Funder, 1995)" (Costa & McCrea, 1999, p.161). Yet later in the same page, the concept of rationality is broadened to all aspects of human's interaction with *reality*: "if our perceptions and judgments were wholly out of touch with reality, we would not have survived as a species" (Costa & McCrea, 1999, p.161). But, the original intension of the text seems to assert that within personality research is that *rationality* is the ability to accurately introspect.

They deem effective introspection to be a requirement for their questionnaire to accurately capture the personality of the individual. They also hold that because humans have survived as a species therefore they are good at introspection. Costa & McCrea (1999) write as if it is not pertinent to explain why this logic applies to humans and not to every other species in the animal kingdom. Does an ostrich accurately introspect that it is open to experiences whereas the earthworm is low in conscientiousness and the microbe is generally agreeable?

Researchers have provided evidence to suggest that our species survived because of perception was adaptive for our ancestral environment (Gilbert, 1998; Haselton, et al., 2009; Volchik, & Zotova, 2013) not because we had an objective view of ourselves. Additionally, others argues that perception is not objective but partially reflects the "costs associated with performing intended actions" to aid decision making (Proffitt, 2006, p. 110). Hills become distorted to appear slanted when wearing a heavy backpack (Bhalla, & Proffitt, 1999). Gaps to jump are reported to be longer when participants' jumping ability is hindered by ankle weights (Lessard, Linkenauger, Proffitt, 2009) or they are fatigued (Johnson, et al., 2021). Individuals who are overweight typically see distances as longer due to the increased effort required (Sugovic, Turk, & Witt, 2016). This distortion of perception can be exploited: making tasks appear easier results in improvements in performance. This occurred in golf (Wood, Vine & Wilson, 2013), weightlifting (Buckingham, Byrne, Paciocco, van Eimeren, & Goodale, 2014) and rifle shooting (Bahmani, Diekfuss, Rostami, Ataee & Ghadiri, 2018). Perception is changed based on changing affordances, and this allows for more appropriate decision making. Critics have called this "Paternalistic vision" (Firestone, 2013, p. 455). Largely, this evidence pushes researchers not only to question if we can perceive the world objective, but also if it is beneficial to do so. There are large schools of thought which now assert that there are biases in human judgement which are profound and extensive (Kahneman, 2017; Thaler, 2017) to the degree that many judgments may be considered irrational. Indeed, there is a pattern of irrationality which is extensively mapped out regarding mistakes that people typically do. Specifically, regarding the definition of rationality from Costa & McCrea (1996) , the question remains if being able to introspect objectively about oneself is that adaptive?

For researchers who do not agree with the assertion that humans are good at introspecting (rationality) there exists a theoretical inconvenience for researching personality. If people are irrational then it seems unlikely that questionnaires will be an effective method to capture who they are as individuals. Research tools that do not rely on the subjective interpretation are available (brain scanners, observational studies, lab studies, etc.) but these methods rarely offer ecological validity data. It may be better to simply establish how a person behaves in their daily life to see how they actually are. Smartphones may offer a method for social psychological researchers free from much major theoretical assumptions as debates for if individuals are effective at introspection. Therefore, this tool may offer an effective method to identify how people actually behaviour and therefore who they are without needing to also assume that individuals are effective introspectors.

7.2 Reliability of psych apps

In 2012, Miller wrote the *Smartphone Manifesto* (Miller, 2012). This represented an initial hypothesis of how useful smartphones could be to psychologists. The formation of this hypothesis was developed only a few years after the launch of the iPhone. Almost, 10 years on the claims contained appear to be reasonable given the evidence available at the time but ultimately immature. To provide some context for the period in which this paper was written, I will quote the first sentence "Smartphones such as the iPhone, Blackberry, and Android" (Miller, 2012, p. 221). The Blackberry has been largely irrelevant on the smartphone marketplace for some time now.

One of Miller's (2012) core claims asserts how useful smartphones could be to psychological researchers. It was argued, the devices could conveniently and consistently deliver high quality and objective records of behaviour[4] and other data from individuals' daily lives. However, substantial issues prevent the devices from becoming the psychological researchers' "go to" tool. Miller did not appreciate that unlike other technologies purpose built for psychological research, smartphones are complex commercial products. Smartphones can be grouped into technologies that are termed as platforms. Creators of the platforms provide

[4] I will only be focusing on smartphones ability to capture behaviour in this conclusion as I have done in this thesis. Smartphones can administer questionnaires, yet this represents a very small contribution to psychological research as little is gained from having questionnaires completed in an ecological setting. Participants will be able to report on their current situation and thus memory limitations should no longer be a factor however responses will remain subjective.

the ability for third parties to deliver content and consumers can access the platform to consume the third parties' content.

Smartphones are not primarily built with the intention that they are used for researchers, and this had profound implications for how reliable psych apps can be or might become. For the smartphone to be commercially viable, there must be an amicable relationship between three parties: the developers of the smartphone operating system, the developers of the apps, and the users of the smartphone. The developers of the operating system must act as authority figures and balance the need of the app developers and smartphone users because design decisions relating to the operating system will impact how the two parties interact. The users' experience must be sufficiently pleasant, or they may switch to another operating system. Equally, app designers need to be rewarded for the consumption of their content. Often, this means extracting details of behaviour which is commercially and psychologically valued such as location, which can also be sold or used for direct advertising (Ketelaar, et al., 2018). However, smartphone operating systems are frequently altered to make such behaviour more difficult to capture due to concerns for the user experience or privacy.

Generally, a smartphone operating system is like a marketplace of content, the operating system developers operate as broker, they set the rules of commerce to reach a fair deal for both content creators and content consumers. The operating system developers noticed that the battery life in Androids was comparatively lower than iPhones despite similar battery types being employed (Android Developer, 2017). The source was discovered that when largely inconsequential (taking a picture or changing location) events were occurring, all apps wanted to extract a record of the behaviour that had occurred and upload the details to a server. A push back on the developer's ability has been on-going. For API level 23 (Android 6.0.0; Android developer, 2019a), the developers of the operating system started implementing doze mode - after phones have not actively been used for a significant amount of time, then almost all background operations would pause until the user opened their phone again. For API level 26 (Android 8.0.0; Android developer, 2019b) there was a limit on the frequency that apps running a background operation (unless they were actively notifying the user of the background operations running; called running a foreground service) could request location update data that they may only get a location update a few times each hour. Additionally, for API level 28 (Android 9.0.0; Android 2019c) the degree that an app is allowed to run in the background would now be directly proportional to the amount of time that the user chooses to actively engage with the app.

The difficulty for the psychological researcher who develops smartphone apps for observational research is that they represent an anomaly for the usual agreements between users and content creators. In Androids, a deal has generally been struck between app developers and smartphone users. It works as follows: the more that a user chooses to engage with an app, the more that app will be allowed to operate in the background. For context, Covid track and trace apps were an exception to the rule, their purpose was deemed sufficiently important to allow unimpeded operating in the background without the user having to choosing to engage with the app (Sabbagh, & Hern, 2020). Without this exception in the rules track-and-trace apps were generally unable to operate effectively. Researchers are generally not constructing apps which are engaging for participants. Psych apps usually have instructions for participant to complete, participants may be contacted to further engage with the app but apps are not generally made to be engaging so that the participant returns to use the app of their own volition in the way popular apps are. I have extolled the value for apps which record participants' behaviour by running a program in the background as it provides the highest quality insight into ecologically valid human behaviour. However, precisely this type of app is mostly likely to be completely stopped from running in the background resulting a complete halting of data being collected. Normally, apps which drain extensive resources by running in the background and are not actively employed by the user would be prevented from wasting resources. Because the smartphone does not identify that there is a distinction from a psych app and regular app then the regular standards are being applied. Henceforth, I will use the term the phenomena of *the operating system type 1 error*.

As will be reviewed later, the psych apps provide highly accurate data which can be assessed for validity in real time. But the operating system type one error (OSTOE) could result in data collection being stopped randomly for different participants. When the smartphone halts background processing may be determined by remaining battery life, amount of usage, the processing power required by other apps. This issue could render an otherwise exciting and potential methodology to be flawed and unreliable. As this issue is so central, I have explored three hypothesized types of solutions: compliance, protesting and micro-psych app.

7.2.1 Compliance solutions

Compliance solutions to the operating system type one error (OSTOE) will involve researchers engaging participants. Fundamentally, this approach accepts the standards that have been placed upon the developers of psych apps and change their design decisions to accommodate the standards. This would work by the participant enjoying engaging with psych apps and therefore the participant routinely engages with the app and the academic can passively monitor the participant behaviour as a result. This solution can be broken down again into collaborative solution and competitive solution.

7.2.1.1 Collaborative-compliance solution

The collaborative-compliance solution involves partnering with industry to study participants who are already engaged with a product. Simply, researchers would request amendments to already existing apps to gather data in the background. Existing research (Johannes, Vuorre, Przybylski, 2021) which has adopted this approach for studying video games demonstrates that it could be effectual but also this effectively alerts readers to the necessary protocols that must be implemented when conducting research with industry. This study involved analysing if the well-being of a sample of players of two different video games is impacted because of their usage. This study is different than many other such studies because the duration of play was objectively established through using the records of Electronic Arts & Nintendo of America. Ultimately, the research found a weak correlation between hours played and increased well-being.

There were two separate games that were involved in the study. The data collection regarding one video game was problematic but not for the other. For the first video game *Plants vs. Zombies: Battle for Neighborville* there was not much control over the data collected by the researchers. They reported that "Electronic Arts (EA) programmed and hosted the survey on Decipher, an online survey platform, and sent invite emails to adult … Electronic Arts then pulled telemetry game data of players" (Johannes, Vuorre, Przybylski, 2021; p.4). Therefore, Electronic Arts would have control of both the survey data and the associated video game usage data. The organisation might well be motivated to return data that suggested video games were good for individuals mental health and change the narrative that had been standing for a lengthy period. There may be significant financial reason to do so. If there was compelling evidence that video games were not harmful to people's wellbeing, but instead were beneficial, it may reduce any reservation that gamers and parents had regarding purchasing more video games. This is exactly what happened, circulation of this study began on the 16th of November 2020 (Yahoo, 2021a), the stock price of the Electronic Arts was $118.60 but the daily closing price linearly ross until the end of 18th of December 2020 the price was $142.61 (An increase in the company valuation of 20%). The same trend occurred for Nintendo (the other company involved in this study) during this time. The stock price of Nintendo was $63.50 on the 16th of November and the daily closing price linearly rose until the 17th of December $82.15 (An increase in the company valuation of 29%; Yahoo, 2021b). Thus, the organisations had opportunity and motive. Such increases in the valuation of the company may well influence the bonus that executives receive and will directly impact the value of assets held by stakeholders.

However, it seems that we can be confident that the data was not manipulated because of how the data collection occurred for the other game *Animal Crossing: New Horizon*. The data collection was controlled by the researchers and the Electronic Arts would only have access to the participants hash number so that the video game usage records could be extracted and therefore no manipulation of the data could effectively occur. We can compare the scores across the two video games for multiple metrics, it appears that the data collection process controlled by the researchers showed that the link between playing video games and wellbeing was more positive than the data returned from the Electronic Arts. If the only data collected in this research was directly controlled by private companies then readers may well doubt that the validity of the data and thus the overall confidence in the findings may be limited.

If a collaborative researcher solution is embraced to combat the OSTOE then there needs to be strong guarantees that the corporate interests do not corrupt the objectivity of the research. From the paper it appears that Electronic Arts in all probability did not alter the data, but there is no guarantee that other companies will not take advantage of researchers and their control of the data. Readers may have greater confidence in the integrity of the findings if protections against corrupting influences had been set up, such as having a third-party oversee data collection analysis, which has been implemented in past research (Ellis, McQueenie, McConnachie, Wilson & Williamson, 2017). Indeed, the research should also adhere to well documented open-source practices wherever possible (Hesse, 2018). Or if corporations make significant influence on how the research is conducted then it essential that such influences are clearly declared

7.2.1.2 *Competitive compliant solutions*

Researchers unable to secure an industry partnership or who are sceptical that any such partnership can foster objective research may be tempted to build their own app to access data. This is a huge technical requirement and puts academics in direct competition with app development companies. There are reasons why an app generated primarily by researchers would struggle to compete for attention in the app market places. First, it is statistically harder to be a popular app developer than a well cited academic, popular classical musician or popular novelist. For content creation generally 80% of success will be shared by 20% of the creators (Taleb, 2007). This applies to 20% of academics receiving 80% of all citations, 80% of classical music played comes from 20% of the composers and 80% of novels read are from 20% of the authors (the rate is worse for non-fiction writers). Conversely it is significantly more difficult to become successful as an app publisher. Since 2014 to 2019 around 80% of all apps (across the App store and Play store) downloaded will be created by just 1% of the app developers (Perez, 2019). Part of this figure is inflated by app publishers buying out successful apps and then republishing the app as their own product. Facebook has done this multiple times (Martinez, 2016). Facebook in a financial quarter will receive a similar amount of app downloads as 100,000 apps developers from the 99% group. Additionally, "More than 95% [of] apps are downloaded by fewer than 1,000 devices" (Liu, et al., 2017, pg. 7). Suggesting if a researcher wishes to develop an app which participants will choose to download for a well powered study (+1000 participants), they may need to develop an app which has the potential to be more popular than 95% of apps (additionally the academics will have to resist the temptation of being bought out by Facebook for millions).

Second, there are many companies which operate internationally and are specialised in app development and marketing and have extensive resources for producing apps. Whereas academics may have difficulty getting resources from funders to build such an engaging app as it could seem peripheral to the fundamental research being conducted. Second, academics attention is typically required to be divided across research, teaching and administration. There may not be the time within an academic community committed to building such a competitive app within a necessary timeline for it to be appealing in the market. Third, it is reasonable to assume that researchers are generally far less experienced at producing apps than app developers and therefore the apps may not be as refined as those on app marketplaces.

However, there are also advantages that a group of academics have relative to app builders. First, the researchers' apps do not have to be commercially viable they only have to be engaging. It will not be necessary on the apps to have adverts, pay walls, or adopt a free-to-play business model. Additionally, this may provide more freedom to the format of the application, large app developers may decide to instead make small variations on historically successful formats. Second, only one application must be successful for a litany of experiments to be tested. If academics build an app which can allow for hosting multiple experiments, then a single successful app could provide a group of academics an enormous amount of high quality ecologically valid data. Whereas multiple unsuccessful apps can represent a sizable loss to an app development company. These points overlap, they both identify the potentially freedom that researchers have in developing apps and how this represents an exciting opportunity.

Note that these collaborative solutions do not provide the psych apps with any more permissions than typical apps. Therefore, any psych apps which embrace this solution will still have to contend with issues like doze mode (Android Developer 2019a) which affects all apps and the related reliability issues.

7.2.2 Protesting solution

Protest solutions would focus upon highlighting to the developers of the operating system that psych apps should be in their own category and not subject to typical regulations of app behaviour. They should be considered an exception, similar to that of the track and trace apps designed to combat Covid-19. However, the episodes surrounding Cambridge Analytica (Wylie, 2019) were most probably a strong lesson for operating system developers not to provide academics with special permissions!

Facebook provided an academic from Cambridge University called Dr Kogan with access to highly restricted permissions to gather data for their research. The permission provided Dr Kogan with an ability to collect data in a very invasive manor. A Facebook user who signed up to his study would allow all of the data on their Facebook account and all of their friends' Facebook accounts to be collected. Thereby an individual member of a friendship group could consent for all their friends' online behaviour to be documented by the researcher.

Unbeknownst to Facebook (Wylie, 2019), Kogan was collaborating with the private company Cambridge Analytica. Kogan's research was funded by Cambridge Analytica and as an agreement both would keep a copy of the data extracted. Kogan would use the data for future research whereas Cambridge Analytica would use the data to build a personality profile for millions of people across the United States. For the people that Cambridge Analytica had a personality profile of, it was thought that they could better target them with messages to dissuade them to vote for members of the Democratic candidate for president (specifically Hillary Clinton). The success of their efforts is unlikely to ever be determined but this was very damaging for Facebook. Cambridge Analytica is a recognisable name and has tarnished the image of Facebook to such an extent that governmental hearings were held to review how Facebook operates (The Committees on the Judiciary and Commerce, Science and Transportation, 2018) although Cambridge University remained largely unscathed. Certainly, Google and Apple would take lessons from Facebook and may be more than hesitant to allow academics generally to have permissions not generally accessible to the public. Additionally, it is not clear how researchers would put pressure upon Apple and Google to add permissions just for academics.

There is a possibility that governments, in the interests of investigating the impact of screen time upon smartphone consumers (UK Parliament, 2019) may pressure developers of smartphone operating system to allow for psych apps to operate in the background unimpeded. This may function as follows, when distributing apps through an app store certain app publishers linked to universities will be certified as developers of psych apps (perhaps if the academics agree not to share the data with individuals or parties with commercial interests). All apps which are designated as products of official psych app developers can request from the user that they run in the background unimpeded. Users would be warned of the impact upon their battery and performance issues that occur as a result. The permission could function in a complex fashion, perhaps the permission would expire after a given point in time, such as the designated conclusion of the experiment, or after a certain amount of elapsed time the participant would have to again approve the permission.

The feasibility of such a solution is questionable. Primarily, placing pressure on such large cooperation's as Google and Apple is difficult. Various governments have not been able to enforce these companies to even pay taxes (Galloway, 2017). Critics have gone so far as to suggest that it may be similarly difficult to force specific changes in the underlying function of the operating systems. Additionally, it is difficult to say that any legislation generated will be future proofed against changes in technology. For instance, Facebook has launched smart glasses (Facebook, 2021), that offer another potential tool for psychological researchers. Would legislation on psych apps apply to smart glasses? If all reprogrammable technology must provide academics with special permissions access, will this stifle innovation for other technology? Finally, pressure on operating systems seems unlikely to result in a repeat of the Cambridge Analytica scandal as the intension of a policy change would be to increase reliability of data gathering methods, and not provide access to additional data. Yet, it is difficult to predict what the impact on operating systems would be and if it could be exploited.

For instance, an academic could fake interest in the blockchain applications of smartphone devices. They could then use smartphones to mine cryptocurrency at the expense of the smartphone users, while they would personally benefit from it.

7.2.3 Mirco-psych app

Smartphone operating systems halt programs operating in the background frequently. One instance, is when the user is not engaging with the app therefore the app is identified as useless to the user and therefore the background operations may be suspended from ever running, described as OSTOE. Other reasons include doze mode (Android, 2019a), where the user has not engaged with the device for a period of time and all apps are halted from running in the background with a few exceptions for apps central to the smartphone's functionality. Given enough time, this will even stop an app from being able to generate notifications. There is a potential hack to overcome all of these issues by exploiting the limitations of enforcement of regulation of app behaviour. The method to do so has been termed the *micro-psych app*. This hopefully represents the next step in the evolution of the Miller's (2012, p 222) "psych app". Micro-psych apps fundamentally describe two different collections of hardware that symbiotically interact - a smartphone and a micro-controller. A micro-controller is a small computer that can be reprogrammed to produce electronic outputs and detect input from sensors. A micro-psych app involves the micro-controller being connected to a smartphone (via a usb or other methods; much of the inspiration for this idea comes from *Chapter III*). This will allow the micro-controller to benefit from the smartphone's battery and allow communication between the two devices. The smartphone contributes:

- A battery source
- A user interface
- Long-term memory storage

- A radio signal (WiFi, SMS) to communicate data to researchers
- People are accustomed to carrying smartphones around

The microcontroller then offers:

- An highly customisable operating system
- A hugely diverse number of sensors to obtain data from
- Backup Long-term memory storage

The fundamental principle of gestalt psychology rings true here - the whole is greater than the sum of the parts. These micro-psych apps do not suffer the limitation of psych-apps and together this may produce a far more reliable system. To demonstrate this I have developed a micro-psych app version of PEG log (described in *Chapter II*). One of the extensive restrictions of the android operating system impacted PEG log - Doze mode (Android 2019a). Doze mode ensures that after a period of inactivity (roughly 15 minutes although this varied across phones) PEG log would be paused to preserve the battery. This would stop the collection of location data.

To overcome the problem of Doze mode a solution was designed and developed (a video demo can be found here - Geyer, 2020a; code can be found here - Geyer, 2020b). However, this solution offers a more heightened control of the data control in general. First, a signal would routinely be sent out of the Arduino Uno (microcontroller; Arudino 2021a) via a USB. When the phone entered doze mode the signal would stop being sent or received. After 5 seconds passed without a signal being received by the microcontroller, the microcontroller would begin logging location data by itself. The microcontroller would use the connected GPS module (ICQUANZX, 2021) and retrieve data every 15 seconds. This data would then be stored by using a SD card module (Diligent, 2021). This was all possible because of specific Arduino packages that handle reading GPS (Hart, 2019). In the demo, the data was uploaded to the android routinely. But in a complete psych app, the data would only be uploaded when the signal resumed being sent.

The initial exploration of this solution was promising, but was greatly limited due to my lack of understanding of circuitry. In the demo, I used an Arudino Uno (68.6 x 53.4mm; Arudino 2021a) due to the output being 3.3v required for some of the sensors. The much smaller Ardunio nano (18 x 45mm; Arudino, 2021b) however could only supply 5v and therefore could not conveniently connect to the GPS module (without smoke appearing, as I discovered).

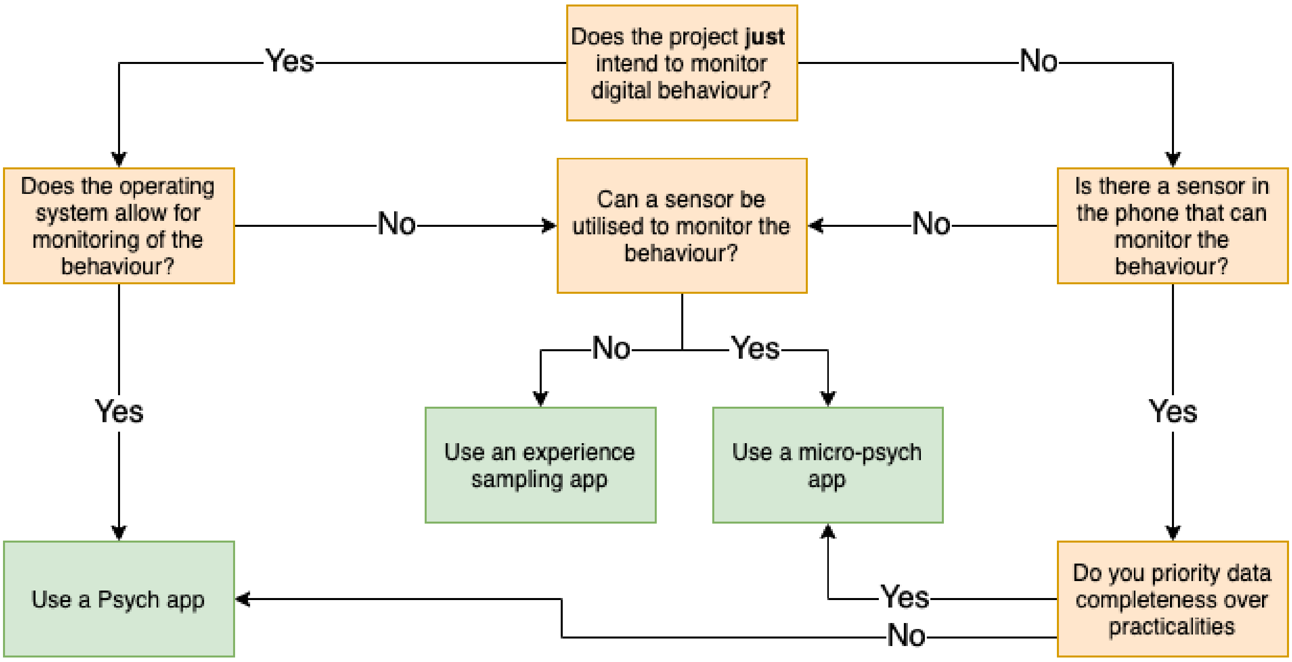
Ultimately, the microcontroller is intended to be packaged into a variety of 3D printed phone cases. The user will simply have to attach the phone case to their device (which provides the USB connection between the microcontroller and the smartphone) and install the appropriate app and the data collection would occur. This is a much more desirable system than simply providing a micro-controller with a battery as participants would have to actively carry the device around with them continually, whereas putting the smartphone in a case means that the researcher is not reliant on the participant continually performing additive behaviour for data collection.

While this proof-of-concept cannot be considered strong evidence that a superior system can be developed to monitor behaviour than the psych app, it provides the rationale for future researchers (myself included) to explore the potential of micro-psych apps. The partnership between smartphone and micro-controller could operate in several ways: the two devices could collect the same data, thus better ensuring the validity of the data. Alternatively, the focus could be on reliability such as in the demo and the devices could trade off allowing operations to proceed for a lengthier period. Another alternative is that the micro-psycho app just runs continually attempting to collect the same data as a smartphone, which offers reliability and validity of data but at a greater cost to the battery. Finally, the devices could work in partnership while collecting separate aspects of data to provide greater insights into the behaviour occurring or the context of the behaviour.

These micro-psych apps provide a final component for the psychological researcher looking to study behaviour. It is the option which is available to researchers when their research would be impacted by the operating system of the smartphone. It is likely that micro-psych apps can work across iPhones and Androids. A flow diagram is provided (see Figure.1) to recommend what method a psychological researcher should use if they wish to passively capture data.

Please note, that this current solution is likely subject to being restricted in the future. Android may well reduce the communication via usb to external devices in future and this could cripple the psych app.

Figure 1: Flow diagram of digital trace solution



Note that this represents the three potential solutions for the OSTOE. However, there are other issues regarding reliability of smartphone apps.

7.3 Validity of Psych apps data

This chapter has thus far extensively reviewed the issues for smartphone reliability and methods to improve upon the consistency for function, especially when logging sensor data. Indeed, there are very significant concerns about researchers' ability to test a participant and find that their data collection has quit after a certain duration. This causes us to question if a representative account of behaviour is being collected. However, the validity of the data being collected is a strength of psych apps.

7.3.1 Corroborating the quality of psych app data

For multiple different psych apps there is an opportunity to corroborate the data in real time. *Peg Log*, described in *Chapter II*, recognised the complexity involved in getting an accurate reading. Location tracking in cities can be difficult because tall buildings could interrupt GPS signals as can heavy clouds (Kijewski-Correa, 2007). However, in areas where multiple Wi-Fi routers are present (very common in cities) they can be used for Wi-Fi Round Trip Time (Android ,2021) to return a smartphone's location with 1-2 meters of accuracy. There are caveats, but generally a psychologist could know how much time a person spends in each room of their house with this technology, given a layout of the house. In more rural settings GPS can deliver accuracy location accuracy up to 5 meters. The *Peg Log* app use Android's available packages (Android, 2021c) to establish the radius where the individual analysing the data can be 68% confident that the individual was inside the radius (it is not publicly known how this is achieved). The validity of the data is being appreciated and taken advantage of across areas in academia. Monitoring the location of members of different communities has allowed for nuance to enter debates about methods of combating prejudice (Dixon et al., 2020). The study allowed for filtering of data which was deemed sufficiently accurate (20 meters).

Similarly, For *TOM* in *Chapter III* the capturing of neurological activation was not conducted but through development there was continual validation of the communication between the app and the brain scanner. When operating properly the app would receive 250 data points per channel each second. The data packets return a timestamp and a data packet id number. Thus, we could guarantee which data packets were being missed because of changes in the id number of the data point. Additionally, we add the ability for researchers to establish how frequently impedance checks would occur to establish high levels of connectivity with the scalp. The participant would also be notified if the connectivity was excessively poor.

Usage Logger in *Chapter IV* did this in a different method. The app could collect real time data of smartphone usage. Simultaneously the Android operating system would collect the same data. Researchers can compare the records from multiple different sources and then were getting similar records of the data. By comparing two accounts (e.g., logbook and extracted data), researchers can have a higher degree of certainty of the validity of the data.

7.3.2 Validating data collection protocol and privacy

For *Usage Logger* in *Chapter IV* did not provide the option to assess the validity of the data in real time. To overcome this, I validated how the app operated through causing actions to occur and reviewing how accurately *Usage Logger* would record the time of that action occurring. The size of the error seemed to be determined by the complexity of the action occurring in the foreground. Thus, the app can be relied upon to detect events occurring on the smartphone (app being opened/closed, notifications being deleted/constructed, screen closing etc.). An important point is that *Usage Logger* cannot verify precisely if an individual is cognitively engaged. Specifically, the app may be on, however it is unclear if the owner is paying attention to it. An app may accidently be left on while the smartphone was charging overnight, this could be discernible by seeing 8 straight hours of consuming the same app (especially if that app functions as a clock or alarm or other such app) but it will not often be obvious when people are and are not paying attention to their smartphone.

There are theoretical options available for ensuring that a researcher knows when an individual is paying attention to their smartphone. For instance, as a system android provides the capability to monitor every button press which occurs on the phone (Android 2021; with extensive caveats). Additionally, when the smartphone is being employed then the self-camera could be employed in combination with facial recognition software to verify that the individual using the smartphone is the intended participant. Such a solution could well raise significant privacy concerns and thus may mean that only a small subset of the population may engage in the study and thus any associated findings of the research may not be generally applicable. Indeed, research should be conducted into what types of smartphone data that people permit to be collected about them in a psychology study. As time passes, concerns regarding privacy may well erode and therefore what a representative sample of the population is willing to have collected about them. There are practicality concerns, such as would the battery be sufficient to monitor the participants' behaviour so much for a meaningful period of time? To avoid such issues, researchers employing psych apps may have to make assumptions about the underlying behaviour of the participants, such as every time the screen is on participants are observing their phone. Or come up with filters that remove times when screens are obviously not been engaged with (e.g. Johannes, Vuorre, Przybylski, 2021).

Indeed, such problems are pervasive amongst psych apps. For instance, the number of WiFi id's identified is used as an indicator of how many people someone typically is around (Wang, & Marsella, 2017). However, there is an assumption here that Bluetooth enabled devices correlate with the number of individuals surrounding the person. However, people could engage with others without it being detected via Bluetooth. Smartphone microphones were also employed to detect the degree to which someone was engaging with others. Monitoring occurrence of speech is separate from detecting meaningful conversation occurring (I am currently reading this thesis aloud and a speech monitoring system may confuse me as currently being social, when I am anything but). More intelligent systems may be constructed to guarantee higher degrees of certainty. For instance, there could be improvements on the conversation detecting system. For instance, machine learning algorithms could distinguish between voices and therefore the research would know exactly when the participant was speaking. Therefore, the research may not confuse a lengthy conversation and the participant being social from the participant listening to a podcast. Through sophisticated methods such issues can be employed to overcome such limitations but the invasiveness of such methods may be extensive.

7.4 Smart psych apps and replication

For many psych apps (none developed in this thesis), a server that collects and orchestrates data collection and data extraction, these may be called smart psych apps. These can offer very valuable methods to collect data and provide a real time ability to monitor data collection. However, the technical skills required for developing such an app are extensive. The app must be built, then integrated with a server, the server must be managed and the server must be protected from physical and cyber-attack. This can be a roadblock for open science. The technical skills required to do this can overwhelm accomplished programmers who have developed multiple apps (this includes myself). Additionally, while many electronic devices can operate as a server, these cheap options expose the server host to multiple security vulnerabilities and therefore expensive commercial options are the only reasonable solution without significant knowledge of cyber security.

For instance, Dixon's and colleagues' (2020) study utilised a psych app which monitors the movements of participants around Belfast. Impressive claims are made regarding the android app (Huck, 2016). They claim that their app recorded the location of participants every 4 seconds. However, I could not find any mention of drop out in data collection, nor of a method that prevents the operating system stopping collecting data of which there are a number (Android 2019a, Android 2019b, Android 2019c). Even though the code is publicly available, because a server was used in this design, there is substantial difficulty involved in testing if the app does not have any data quality problems. For open and transparent research practices, it is much easier for researchers to employ serverless solutions and utilise software that can be directly downloaded from app stores than relying on others setting up their own servers to replicate results.

7.5 General thesis limitations and challenges

The thesis has had several limitations that were associated with the research. For example, a study into the location of participants over a two-week period was partially conducted but the study was abandoned after significant difficulty recruiting participants. Understandably, participants were reluctant to allow the researchers to continually monitor their location for a two-week period. Other problems included issues with an early version of the location recording app. Additionally, multiple participants were not transparent about their future plans regarding leaving the local area as staying in the local area was a part of the eligibility requirements. Ultimately, the study was shut down after few people participated and many of those who did participate would experience issues with the app (a very early version of PEG log) or would not follow experimental procedure. From this experience, a next step could be conducting research with location sensing with superior software and a better idea regarding the recruitment practices that would need to take place to ensure greater participation (also after national lockdowns are no longer in effect).

Another limitation is that this study of digital traces only focused on one of the two primary smartphone operating systems - Android. The iPhone makes up a significant proportion of the international smartphone market and is the best-selling product of all time (if you consider all versions of the device, the same product; Merchant 2017). iPhone users also as a group have a different personality than android users (Shaw, Ellis, Kendrick, Ziegler, & Wiseman, 2016). But, the iPhone share is typically owned by western, educated, industrialized, rich and democratic individuals who are overly represented in the psychological literature (WEIRD; Henrich, Heine, Norenzayan, 2010). There is, however, significant difficulty involved with developing apps for iPhones and publishing them to the app store (Apple, 2021). Many of the options available on android are not on iPhone such as retrieving highly detailed usage monitoring records. While we did develop a location sensing device on the iPhone it could not rival the android version (Peg Log; chapter II). Background logging could not restart upon the phone restarting. Therefore, any participant who would turn off their phone or remove the app from the list of apps allowed to run in the background (easily done in an iPhone) would stop data collection. Once the app was mostly developed, it became apparent that we were limited to install apps onto a handful of participants' iPhones locally by the Apple developer platform. The only other option for installing the app on participants' phone was publishing it on the app store and that would be very difficult due to Apple's secret and rigid criteria of what is allowed to be published on the store. As a result of these multiple difficulties the project was abandoned.

7.6 Thesis Conclusion

Psychology has struggled to effectively and accurately monitor ecologically valid behaviour since its inception. I have provided three different applications that deliver such methodologies. One of these applications was used and highlighted issues of previously employed methodologies and brought into question the link between negative health/mental health and smartphone usage. Generally, psych apps can deliver highly accurate data from sensors and associated databases but its collection can be unreliable. There are three broad categories of solutions that can be employed to over come this issue. How this issue is overcome will determine if smartphones are considered a more standardised method to conduct psychological research in the future. Reflecting a diverse history, there will always be appetite for methodological innovation across psychological science.

References

Android (2021) Wifi location: ranging with RTT [Website]. Retrieved from: ‘https://developer.android.com/guide/topics/connectivity/wifi-rtt’

Android Developer (2017) Background Check and Other Insights into the Android Operating System Framework (Google I/O '17). Retrieved from: https://www.youtube.com/watch?v=hbLAzwhBjFE

Android Developer (2021a) *Build more accessible apps* [Website]. Retrieved from ‘https://developer.android.com/guide/topics/ui/accessibility’

Android Developer (2021b) *Android Architecture* [Website]. Retrieved from ‘https://developer.android.com/guide/topics/ui/accessibility’

Android Developer (2021c) *FuseLocationProviderClient* [Website]. Retrieved from: ' https://developers.google.com/android/reference/com/google/android/gms/location/FusedLocationProviderClient'

Android Developer [Software] (2019a) Optimize for Doze and App Standby. Retrieved from https://developer.android.com/training/monitoring-device-state/doze-standby

Android Developer [Software] (2019b) Background Location Limits. Retrieved from https://developer.android.com/about/versions/oreo/background-location-limits

Android Developer [Software] (2019c) App Standby Buckets. Retrieved from: https://developer.android.com/topic/performance/appstandby

Apple (2021) App Store [Software]. Retrieved from: ‘https://developer.android.com/topic/libraries/architecture’

Arudino (2021) Arudino Uno [Hardware]. Retrieved from: ‘https://store.arduino.cc/arduino-uno-rev3’

Aurdino (2021b) Arduino Nano [Hardware].Retrieved from: ‘https://www.arduino.cc/en/pmwiki.php?n=Main/ArduinoBoardNano’

Bahmani, M., Diekfuss, J. A., Rostami, R., Ataee, N., & Ghadiri, F. (2018). Visual illusions affect motor performance, but not learning in highly skilled shooters. Journal of Motor Learning and Development, 6(2), 220-233.

Bhalla, M., & Proffitt, D. R. (1999). Visual–motor recalibration in geographical slant perception. Journal of experimental psychology: Human perception and performance, 25(4), 1076.

Buckingham, G., Byrne, C. M., Paciocco, J., van Eimeren, L., & Goodale, M. A. (2014). Weightlifting exercise and the size–weight illusion. Attention, Perception, & Psychophysics, 76(2), 452-459.

Costa, P. T., & McCrae, R. R. (1999). A five-factor theory of personality. The Five-Factor Model of Personality: Theoretical Perspectives, 2, 51-87.

Diligent (2021) Development Kit Pmod MicroSD Card Slot for use with Store and Access On System Board [Hardware]. Retrieved from: ‘https://uk.rs-online.com/web/p/memory-development-tools/1840482/?cm_mmc=UK-PLA-DS3A-_-google-_-CSS_UK_EN_Raspberry_Pi_%26_Arduino_%26_Development_Tools_Whoop-_-Memory+Development+Tools_Whoop-_-1840482&matchtype=&pla-301255893531&gclid=EAIaIQobChMIrubBtLKl7wIVELrtCh3vbwruEAQYAyABEgIxEvD_BwE&gclsrc=aw.ds’

Dixon, J., Tredoux, C., Davies, G., Huck, J., Hocking, B., Sturgeon, B., ... & Bryan, D. (2020). Parallel lives: Intergroup contact, threat, and the segregation of everyday activity spaces. Journal of Personality and Social Psychology, 118(3), 457.

Dumont, F. (2010). A history of personality psychology: Theory, science, and research from Hellenism to the twenty-first century. Cambridge University Press.

Ellis, D. A., McQueenie, R., McConnachie, A., Wilson, P., & Williamson, A. E. (2017). Demographic and practice factors predicting repeated non-attendance in primary care: a national retrospective cohort analysis. The Lancet Public Health, 2(12), e551-e559.

Emre, M. (2018). The personality brokers: The strange history of Myers-Briggs and the birth of personality testing. Random House Canada.

Facebook (2021) Ray-Ban and Facebook introduce Ray-Ban Stories, first-generation smart glasses. Tech@Facebook [website]. Retrieved from: 'https://tech.fb.com/ray-ban-and-facebook-introduce-ray-ban-stories-first-generation-smart-glasses/'

Firestone, C. (2013). How "paternalistic" is spatial perception? Why wearing a heavy backpack doesn't—and couldn't—make hills look steeper. Perspectives on Psychological Science, 8(4), 455-473.

Galloway, S. (2017). The four: the hidden DNA of Amazon, Apple, Facebook and Google. Random House.

Geyer, K., (2020a). Micro-psych app: Retrieved from: https://www.youtube.com/watch?v=7-YXY_ySL_8&ab_channel=KristofferGeyer%60

Geyer, K., (2020b). Micro-psych app: Retrieved from: https://github.com/kris-geyer/micro-psych-app

Gilbert, P. (1998). The evolved basis and adaptive functions of cognitive distortions. British Journal of Medical Psychology, 71(4), 447-463.

Hart, H. (2019) TinyGPS++ [Software]. Retrieved from: 'https://github.com/mikalhart/TinyGPSPlus'

Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. Social Cognition, 27(5), 733-763.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. Nature, 466(7302), 29-29.

Hesse, B. W. (2018). Can psychology walk the walk of open science?. American Psychologist, 73(2), 126.

UK Parliament (2019). Impact of social media and screen-use on young people’s health, Fourteenth Report of Session 2017-19, Retrieved from: 'https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf'

Huck, J. (2016) Belfast Pathways GPS Tracker for Android. Retrieved from: https://github.com/jonnyhuck/bmp-pathways-app

ICQUANZX (2021) ICQUANZX GY-NEO6MV2 NEO-6M GPS Flight Controller Module 3V-5V with Super Strong Ceramic Antenna for Arduino EEPROM APM 2.5. retrieved from ‘https://www.amazon.co.uk/ICQUANZX-GY-NEO6MV2-Controller-Ceramic-Antenna/dp/B088LR3488/ref=asc_df_B088LR3488/?tag=googshopuk-21&linkCode=df0&hvadid=463207589700&hvpos=&hvnetw=g&hvrand=5966689899117451231&hvpone=&hvptwo=&hvqmt=&hvdev=c&hvdvcmdl=&hvlocint=&hvlocphy=1007043&hvtargid=pla-946390196873&psc=1’

Johannes, N., Vuorre, M., & Przybylski, A. K. (2021). Video game play is positively correlated with well-being. Royal Society open science, 8(2), 202049.

Johnson, C. D., Eagle, S. R., Nindl, B. C., Lovalekar, M. T., Flanagan, S. D., Pepping, G. J., & Connaboy, C. (2021). The effects of fatiguing exercise and load carriage on the perception and initiation of movement. European journal of sport science, 21(1), 36-44.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Ketelaar, P. E., Bernritter, S. F., van Woudenberg, T. J., Rozendaal, E., Konig, R. P., Hühn, A. E., ... & Janssen, L. (2018). “Opening” location-based mobile ads: How openness and location congruency of location-based ads weaken negative effects of intrusiveness on brand choice. Journal of Business Research, 91, 277-285.

Kijewski-Correa, T., & Kochly, M. (2007). Monitoring the wind-induced response of tall buildings: GPS performance and the issue of multipath effects. Journal of Wind Engineering and Industrial Aerodynamics, 95(9-11), 1176-1198.

Lessard, D. A., Linkenauger, S. A., & Proffitt, D. R. (2009). Look before you leap: Jumping ability affects distance perception. Perception, 38(12), 1863-1866.

Liu, X., Ai, W., Li, H., Tang, J., Huang, G., Feng, F., & Mei, Q. (2017). Deriving user preferences of mobile apps from their management activities. ACM Transactions on Information Systems (TOIS), 35(4), 1-32.

Martínez, A. G. (2016). Chaos monkeys: Inside the Silicon Valley money machine. Random House.

McCrae, R. R., & Costa Jr, P. T. (2008). The five-factor theory of personality.

Merchant, B. (2017). The one device: The secret history of the iPhone. Hachette UK. WEIRD

Miller, G. (2012). The smartphone psychology manifesto. Perspectives on psychological science, 7(3), 221-237.

Perez, S. (2019) The top 1% of app store publishers drive 80% of new downloads. *Tech Cruch*. Retrieved from: https://techcrunch.com/2019/11/21/the-top-1-of-app-store-publishers-drive-80-of-new-downloads/

Sabbagh, D., & Hern, A. (2020). UK abandons contact-tracing app for Apple and Google model. The Guardian. Retrieved from: https://www.theguardian.com/world/2020/jun/18/uk-poised-to-abandon-coronavirus-app-in-favour-of-apple-and-google-models

Shaw, H., Ellis, D. A., Kendrick, L. R., Ziegler, F., & Wiseman, R. (2016). Predicting smartphone operating system from personality and individual differences. Cyberpsychology, Behavior, and Social Networking, 19(12), 727-732.

Simon, H. (1957a). A behavioral model of rational choice. Models of man, social and rational: Mathematical essays on rational human behavior in a social setting, 241-260.

Simon, H. A. (1957b). Models of man: Social and rational. Oxford, UK: Wiley.

Sugovic, M., Turk, P., & Witt, J. K. (2016). Perceived distance and obesity: It's what you weigh, not what you think. Acta psychologica, 165, 1-8.

Taleb, N. N. (2007). The black swan: The impact of the highly improbable (Vol. 2). Random house.

Taleb, N. N. (2020). Skin in the game: Hidden asymmetries in daily life. Random House Trade Paperbacks.

Thaler, R. H. (2017). Behavioral economics. Journal of Political Economy, 125(6), 1799-1805.

The Committees on the Judiciary and Commerce, Science and Transportation. 197st Cong. (2018) (testimony of Mark Zuckerberg).

The Committees on the Judiciary and Commerce, Science and Transportation. 197st Cong. (2018) (testimony of Mark Zuckerberg).

Volchik, V., & Zotova, T. (2013). Adaptive rationality in the evolutionary and behavioral perspectives. Middle East Journal Of Scientific Research, 17(10), 1488-1497.

Wang, H., & Marsella, S. (2017, October). Assessing personality through objective behavioral sensing. In 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 131-137). IEEE.

Wood, G., Vine, S. J., & Wilson, M. R. (2013). The impact of visual illusions on perception, action planning, and motor performance. Attention, Perception, & Psychophysics, 75(5), 830-834.

Wylie, C. (2019). Mindf* ck: Inside Cambridge Analytica's plot to break the world. Profile Books.

Yahoo (2021a) Electronic Arts Inc. (EA) [Website], Yahoo finance. Retrieved from: ' https://finance.yahoo.com/quote/EA/history?p=EA'

Yahoo (2021a) Nintendo Co., Ltd. (NTDOY) [Website], Yahoo finance. Retrieved from: ' https://finance.yahoo.com/quote/NTDOY/history?p=NTDOY'