University of Basel

Faculty of Psychology

Bachelor's thesis presented to the Department of Psychology of the University of Basel for the degree of Bachelor of Science in Psychology

Could Social Media Platforms Provide New Ways to Measure Personality? – A Small Sample Scoping Review

Author: Boas Loeb, MLaw, BA

Immatriculation number: 07-059-587

Correspondence email: boas.loeb@stud.unibas.ch

Examiner: M.Sc. Sharon Steinemann

Supervisor: Prof. Dr. Klaus Opwis

Center for Cognitive Psychology and Methodology

Submission: 18. April 2017

Content

Abstract

Recent research shows great promise for predicting personality from social media data. This preliminary small sample scoping review provides an idea on the possibilities that social media data and social media platforms could provide for measuring personality. The review suggests that scientifically designed online environments or applications could provide interesting possibilities for collecting and analysing personal, social and mass-behavioural data. Furthermore, social media users interest to self-present align well with the interests of personality researchers, which suggests valuable motivational resources. A theoretical framework of these possibilities is provided as well as experiences regarding the small sample scoping review method used in this study.

Keywords

Social Media, Personality Assessment, Data collection, Big Five, Scoping review, Self-Presentation

1. Introduction

Social media is well established as a data source in different scientific disciplines, but a very young research topic and not a very coherent one (Conway & O'Connor, 2016). Different perspectives and disciplines use different datasets, data types and methods to answer different research questions (Weller & Kinder-Kurlanda, 2015). Recent studies show promise for predicting personality, using millions of words, phrases and topics, from tens of thousands of social media users (Park et al., 2015; Schwartz et al., 2013). There are several good reasons why social media might provide new and meaningful ways for measuring and understanding personality. First, social media users disclose a lot of information on social media about their true selves and not just idealized versions (Back et al., 2010). Social media users tend to disclose a lot of personal information, which, due to its digital nature, can be easily recorded (Park et al., 2015). This allows researchers to observe and measure natural behaviour of users in their environment and to capture real communication between friends and acquaintances (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015). It allows essentially ongoing experimental sampling of people's real social lives and behaviour (Park et al., 2015). The far-spread use of social media platforms enables Big Data analysis, which facilitates the discovery of patterns that might not be seen in smaller samples. Social media populations furthermore provide data well beyond the W.E.I.R.D samples (Western, Educated, Industrialised, Rich, Democratic; Henrich, Heine, & Norenzayan, 2010) used for many psychological studies, and the high statistical power of Big Data analyses might address the replicability crisis in psychological sciences (Kosinski, Wang, Lakkaraju, & Leskovec, 2016). This provides new opportunities for studying aspects of human behaviour that were previously very difficult - if at all possible - to assess (Wilson, Gosling, & Graham, 2012). There are several drawbacks though: handling big amounts of data and efficient usage of online environments requires computational and statistical skills, which are not necessarily common knowledge (Bello-Orgaz, Jung, and Camacho, 2016). Since researchers may lack the necessary skills, there is the danger of tasks being ceded to computer scientists, which in term may lack the theoretical background (Kosinski et al., 2016). Multidisciplinary collaborations are therefore required to access the full potential of social media platforms for science (Schwartz et al., 2013). Multidisciplinary collaborations though are time-consuming and require a major commitment of researchers from different disciplines. Furthermore, social media (big) datasets are not readily available to researchers, as social media platforms are being developed and maintained by private corporations whose financial interests might collide with those of the users as well as the researchers. This leads to several ethical concerns, which might aggravate in the future. Mittelstadt and Floridi (2016) meta-analytically identified key areas of ethical concern to be *informed consent*, *privacy and ownership of the data.* Of further concern are the *'Big Data Divides'* between those who have and those who lack the necessary resources or connections to analyse large social media datasets (Mittelstadt & Floridi, 2016). Regarding these concerns, a clear distinction between commercial and academic uses of social media data is suggested (Mittelstadt & Floridi, 2016). Social media platforms provide as well interesting and ethical possibilities to collect data without the help of commercial social media platforms. Huang et al. (2015) suggests three ways for collecting social media data: a) by directly retrieving data shared on social network websites, b) by asking participants about their behavior and c) through deployed applications. Social media data can be extracted directly from available online records, for example, using so called 'crawlers' (as done for example by Maria Balmaceda, Schiaffino, & Godoy, 2014). Such an approach though should be considered very carefully from an ethical point of view, due to its lack of informed consent, dangers for privacy and the lack of clarity regarding the ownership of the data. Collaborating with social media providers might also be problematic, since it could blur the distinction between commercial and academic data uses. The second option, asking social media users to provide certain data seems to be a more ethical method. This method has been discussed as 'crowdsourcing', for example by using Amazon Mechanical Turk to recruit participants (Kosinski, Bachrach, Kasneci, Van-Gael, & Graepel, 2012). Amazon Mechanical Turk is an online labor market that assists researchers (or other requesters) in recruiting and compensating workers for a variety of tasks and has been widely used in social science (Cheung, Burns, Sinclair, & Sliter, 2016). The third solution is to compute specific applications or online environments which collect the required data. One such successful example is the MyPersonality App from David J. Stillwell and Michal Kosinski, which has been deployed in June 2007. Over 6 million users completed the most popular questionnaire, a short version of the NEO Personality Inventory, and almost half of the users allowed the researchers to anonymously record information from their Facebook profiles, without any financial incentives (Stillwell & Kosinski, 2004). This dataset was continuously used for different research projects and provided the data, for example, for the studies mentioned by Park et al. (2015) and Schwartz et al. (2013). The potential of this dataset has not been exhausted and the dataset is freely available for researchers at http://www.mypersonality.org/. Such methods provide free access to data, which further limits the divide (Mittelstadt & Floridi, 2016). The example of MyPersonality shows, that scientific online applications can make use of social media platforms, without depending on them. Computing scientific applications or other scientific online environments have the potential to be an important source for collecting personal data and, if reliable methods have been developed, measure personality for academic purposes. As stated before, developing such a platform or application would require a multidisciplinary collaboration which is time-intensive and requires major comitments. Before undertaking such a project, a preliminary study was thought to be useful to determine the value of such collaborations and provide the necessary background for success.

2. Methods

The main research question, whether social media platforms could provide new ways for measuring personality, is a very broad question and as such makes it difficult to choose an adequate methodology. Unfortunately, to stay within the scope of this analysis, no empirical study could be conducted. Confirmation bias is a frequent problem in research, but especially relevant in preliminary studies (since researchers might already have a theory or concept in mind they wish to prove) and literature reviews (since there are no objective measures, which literature has to be included). Systematic literature reviews and meta-analyses are two kind of approaches that are less susceptible to confirmation bias, since they aim to include *all* available literature. However, systematic reviews and meta-analyses are very time-intensive studies and are usually conducted by teams of experienced researchers who focus on much narrower questions. For such a broad research question, a different form of systematic review, a so-called scoping review (also called ‘systematic mapping studies’ or ‘scoping studies’) is recommended (Keele, 2007). Scoping reviews are similar to systematic literature reviews, except that they employ broader inclusion criteria and are intended to map out topics rather than synthesize study results (Keele, 2007). They are also used for the purpose of identifying gaps in research literature (Arksey & O’Malley, 2005) and are useful as preliminary studies to determine whether future work, e.g. full systematic reviews, would prove of worth. Scoping reviews are an approach to review literature which has received little attention in the research methods literature (Arksey & O’Malley, 2005). Furthermore, the scoping reviews that exist vary widely in terms of intent, procedure and methods (Davis, Drey, & Gould, 2009; Pham et al., 2014). Guidelines on how such a systematic scoping review should be conducted are not well established, especially for preliminary studies with the given limitations. Arksey and O’Malley (2005) proposed a framework, which has been further developed by Levac, Colquhoun, and O’Brien (2010), Daudt, van Mossel, and Scott (2013) and O’Brien et al. (2016). Since there is still little experience with this approach, researchers are encouraged to experiment and provide additional experiences (Levac, Colquhoun, and O’Brien, 2010). For the following purposes, the framework proposed by Arksey & O’Malley (2005) and the recommendations by Levac et al. (2010) and Daudt et al. (2013) have been used and adapted to fit the specific needs for this study (Table. 1).

Table 1. Framework adapted from Arksey and O’Malley (2005)

Stage 0: Intuitive literature review to broaden the understanding, generate hypothesis, identify possible research gaps.
Stage 1: Identifying detailed research questions and the relevant topics, weighting them according to suspected importance.
Stage 2: Identifying relevant studies through searches including a set amount of results, depending on set importance.
Stage 3: Study selection: Exclude unavailable or doubled data. If necessary, include literature on different levels.
Stage 4: Charting the data: Extract general data first. Continuously adapt extraction forms for specific information.
Stage 5: Collating, summarizing, and reporting the results: 1. Analyse data, 2. create a theoretical framework or structure encompassing all extracted information, 3. Report the findings, 4. Answer detailed research questions and add conclusion, limitations and discussion
Stage 6: Consultation (optional)**:** Consultation with key stakeholders may provide additional sources of information and offer different perspectives on the data collected.

Arksey and O’Malley (2005) elaborate four main reasons why a scoping study might be undertaken: (a) to examine the extent, range and nature of research activity, (b) to determine the value of undertaking a full systematic review, (c) to summarise and disseminate research findings and (d) to identify research gaps. This study aims to determine the value of undertaking a multidisciplinary collaboration, but focuses as well on identifying research gaps where such a collaboration could prove especially useful. It further aims at providing an overview or at least an adequate idea of the research landscape and finally, to counteract the confirmation bias expected in such a preliminary literature review by using some sort of objective measure. The first stage according to Arksey and O'Malley (2005) seeks to identify detailed research questions. But identifying research questions requires a deep understanding of the subject and is a creative task, requiring to hypothesize about relevant topics and questions. Therefore, an additional stage 0 has been added, which includes a broad and intuitive literature research. This research was required to gain the additional background knowledge, to identify topics that are relevant and deepen the methodological understanding. For this stage, a separate research question was formulated, pointing toward identifying potentially interesting applications. Specifically, it was asked whether social media platforms could provide *new* ways to measure personality in comparison to the predominant way of measuring personality, the Big Five (Matz, Chan, & Kosinski, 2016). The first research question was therefore defined as follows: 'What are potentially relevant limitations of the predominant way of measuring personality?'. To answer this question, a wide aray of search queries on different search engines (Google Scholar, PubMed, Scopus) as well as manual searches were conducted, scanning well over a hundred different results at different levels of depth for relevant information. This first research question aimed at identifying areas in which social media platforms could provide new possibilties and was therefore considered a creative question. Searches and inclusion of the results were conducted intuitive rather than systematic and were not documented, since this was thought to disrupt and limit the creative process. The results were then used to develop more precise research questions for future research and identify relevant topics (*Stage 1*). The thinking which led to the detailed research questions is listed in the following section. Next, a main research question which was found to fit the goals was phrased as follows: 'What are the possibilities of social media data for predicting personality discussed in literature?'. This research question aims to provide an overview of recent research activity and determine the value of undertaking a multidisciplinary collaboration. Appendix A provides an overview over the different research question, the relevant topics and search queries.

Stage 2 of the framework focuses on identifying the relevant studies and involves searching for research evidence, using different sources (Arksey & O'Malley, 2005). One important aspect of this stage is that practical issues related to time, funding, and access to resources often require researchers to consider the balance between feasibility, breadth, and comprehensiveness (Levac et al., 2010). Levac et al. (2010) therefore recommend that research questions and purpose should guide decision-making around the scope of the study. If limiting the scope of the research is unavoidable, decisions should be justified and the potential limitations of the study acknowledged. Daudt et al. (2013) proposed for this stage to build both a multidisciplinary and inter-professional team and to include someone experienced with scoping studies if possible. Since this study had several limitations regarding time and personal resources, it was therefore concluded that the scope of this study had to be seriously limited. The possibility, that relevant literature could be missed is reported as the most common limitation of scoping reviews and some argue that it might not be realistic for scoping reviews to retrieve and include all relevant literature due to the broader focus (Pham et al., 2014). The lack of clear boundaries, left some researchers with overwhelming amount of data, challenges the feasibility especially, since scoping reviews often took longer than originally anticipated (O'Brien et al., 2016). On the other hand, the ability to provide an overview over a broad research question is seen as one of the core strenths of scoping reviews (O'Brien et al., 2016). With the given limitations, especially regarding the deadline, this review chose to limit the number of search results, which were included. For the main research question, the first 20 results from a Google Scholar search were thought to be a sample for relevant literature that is being discussed. Additional searches of less important topics included only the first 5 results search results. The limitation on the number of search results made use of the ranking algorithms provided by Google Scholar. The intrinsic problem with this sort of limitation is, that Google does not, in contrast to other scientific search engines, e.g. Scopus, disclose the criteria for its ranking algorithm. This makes it very difficult to assess whether the search conducted did really list relevant literature or whether the sample could be representative for the research landscape. On the other hand, this lack of disclosure might make ranking less vulnerable to academic search engine optimisation (Martin-Martin, Orduna-Malea, Harzing, & López-Cózar, 2017). Through reverse engineering, Beel & Gipp (2009) found citations as well as the title to be of particular importance for Google Scholars ranking, though other - not identified - aspects play an important role as well. A study done by Martin-Martin et al. (2017) found, that Google Scholar is able to reliably identify the most highly-cited academic documents which suggests Google Scholar to be a valuable tool to identify influential scientific work. Google Scholar is as well widely used, especially as a source for unpublished research in systematic literature reviews (Haddaway, Collins, Coughlin, & Kirk, 2015). Systematic reviews typically screen the first 50 – 100 Google Scholar search records, though it has been proposed that the focus should be expanded to the first 300 Results (Haddaway, Collins, Coughlin, & Kirk, 2015). For the following purposes, it was neither possible nor intended to adequately picture *all* of the literature available. Instead, the main objective was to create an idea of the relevant literature and research landscape, while counteracting the expected confirmation bias. Whether such an approach is or could be adequate or even useful is further discussed in the limitation section.

Stage 3 regards the selection of the studies. Ar ksey & O'Malley (2005) propose to use inclusion/exclusion criteria developed post hoc after familiarity with the literature is established, whereby a team approach is suggested but not imperative. Since the scope of the study was already very limited and the ranking of the literature was thought to be of importance, almost no studies were fully excluded. No paper was for example excluded due to its lack of relevance, since this could potentially nullify the objectivity that was aimed at. Some literature though was not or only partially available and certain publications were found two times in a search result, e.g. a conference paper as well as the published paper. Such results were still included regarding the analysis of the literature found, since the ranking in the search result was assumed to imply their importance. Other papers focused on a marketing perspective. Due to the lack of an academic background in this discipline, they were included only on a superficial level. This led to three different levels of inclusion: a) literature included only for literature analysis, b) literature included only on a superficial level and c) fully included literature.

Stage 4 included charting the data. Arksey and O'Malley (2005) describe this step as collecting data according to key issues and themes, whereby two main categories of data are suggested: general information about the study and specific information related to the research question. (Levac et al.,

2010) and (Daudt et al., 2013) suggest that charting should be considered an iterative process in which researchers continually extract data and update the data-charting form, where trial charting should be conducted. In accordance with these suggestions, first general data, as well as background information, were extracted in a special form. Different additional charting forms were developed and updated many times regarding the specific information related to the research questions. It was confirmed that data forms had to be adapted many times during the process and need to be specifically designed for the relevant research question. The final data extraction forms used in this study are available from the author, if requested.

Stage 5 includes collating, summarizing, and reporting the results: Arksey and O'Malley (2005) propose that this steps includes a descriptive numerical summary related to the general information collected and a thematic construction of the specific information collected. (Levac et al., 2010) propose three distinct steps: a) analysis (including descriptive numerical summary analysis and qualitative thematic analysis), b) reporting the results and producing the outcome that refers to the overall purpose or research question and c) considering the meaning of the findings as they relate to the overall study purpose; discussing implications for future research, practice and policy. These three steps proved to be useful, but challenging. The variables which were of interest regarding the literature analysis, for example, could only be established after the meaning of the findings and how they relate to the overall study purpose were considered (c). For reporting the results (b) different methods were applied to put the available data into context and to present it in a coherent way. First, different keywords or topics were added to the general information, which were then ordered and put into a structure. This information was then included in an existing theoretical framework for social media research, provided by Ngai 2015, though several adaptations had to be made. The final step according to Arksey & O'Malley (2005), *Stage 6,* consultation with key stakeholders, would have exceeded the scope of this thesis and had to be omitted. The following section provides the results of the first phase of literature research, regarding the limitations of the predominant way of measuring personality.

3. Limitations of the Predominant Method of Measuring Personality

Allport (1937) defined personality as something that includes 'all of the attributes, qualities, and characteristics that distinguish the behavior, thoughts, and feelings of individuals'. Personality is a very broad term, which makes it a particularly difficult to measure. There is furthermore no all-encompassing theory of personality, which allows no theory-based development of a measure (Rust, Golombok, Kosinski, & Stillwell, 2014). Personality measures are therefore pragmatic approximations and standardized frameworks but do not account for the full complexity of individual behaviour (Matz, Chan, & Kosinski, 2016). Most personality measures are based on the lexical hypothesis, first introduced by Sir Francis Galton, stating that every perceivable aspect of personality is encoded in the natural language (Matz, Chan, & Kosinski, 2016). Allport and Odbert combed through 17'953 words of the English dictionary identifying words, that could be used to describe others (Matz, Chan, & Kosinski, 2016). They identified four word categories: (1) personality traits, (2) present states, attitudes, emotions and moods, (3) social evaluations and (4) others (Matz, Chan, & Kosinski, 2016).

Since they included only personality traits in their final taxonomy, the Big Five are understood as a trait-focused approach (Wilson, Thompson, & Vazire, 2016). The factor analysis of these traits resulted in the well-known Big Five dimensions of personality (Openness, Agreeableness, Consciousness, Neuroticism, Extraversion), which have since then shown to be an indispensable and reliable tool for measuring personality (Matz, Chan, & Kosinski, 2016). As a trait-focused approach, it has long been debated whether the five-factor model is useful to describe variations in psychological states over time as well (Borkenau & Ostendorf, 1998). This 'person-situation' debate has been additionally sparked through longitudinal research, showing traits not to be as stable as they were once thought to be, both across occasions and over situations (Matz, Chan, & Kosinski, 2016). Also, the prediction of future behaviours based on current traits turned out to be less than impressive (Lucas & Donnellan, 2009). Consequently, social psychologists argued that what people think, feel and behave at any given moment is largely dependent on the situation. In contrary, personality researchers believed that thoughts, feelings, and behaviors are relatively consistent (Lucas & Donnellan, 2009). This debate could be mostly resolved through the development of the integrative trait-state models (Hamaker, Nesselroade, & Molenaar, 2007). According to the integrative trait-state model, individuals might, for example, be more extroverted in certain situations, (e.g. at a party), but individuals with higher extroversion would still be more extroverted in the same situation than a less extroverted individual (Matz, Chan, & Kosinski, 2016). Recent studies suggest that the within-person variability in personality states cannot be accounted for entirely by fluctuations in affect (Wilson et al., 2016) and that velocity, meaning cognitive goals, may have an important influence on within-person variability of personality traits (Wilt, Bleidorn, & Revelle, 2016). These findings suggest that the Big Five may at least be limited in their capability to picture within-person variability. New technologies provide new possibilities to collect and analyse intensive longitudinal data on personality states (Wilson et al., 2016). This should account for social media sites as well, which therefore could provide new insights into personality states and situational factors. This leads to the first research question, *R1: Could Social media platforms provide new ways for measuring relevant aspects of the within-person variability?*

Further possibilities might be found in the lexical base, from which the big five are derived. There is an ongoing debate, since some researchers do not see the lexical approach as a valid method of scientific investigation due to different biases, e.g. a pro social behaviour bias (Trofimova, 2014). Additionally, the lexical base of the Big Five is said to be invalid, because it uses descriptions of language rather than actual language behaviour (Trofimova, 2014). Furthermore, it has been questioned whether the Big Five factors are stable over different cultures and languages. Some argue that other number of factors, for example six, as proposed in the Hexaco Six Factor Model, might work better across different languages and cultures (Ashton, Lee, & de Vries, 2014). It has therefore been suggested that the widespread adoption of five-dimensional personality models was premature (Ashton, Lee, & de Vries, 2014). Other findings show the Big Five to be stable across cultures as well as instruments and observers (Matz, Chan, & Kosinski, 2016). In regards to social media, the 'internet language' might again differ substantially from normal language use and provide interesting possibilities due to it being shaped by a multicultural population. This leads to the second research question, *R2: Could social media provide new ways regarding the lexical base?*

Furthermore, a limitation can be seen in the abstraction or broadness of the Big Five personality dimensions. Personality traits can be defined with different degrees of conceptual breadth. A broadly defined trait (e.g. conscientiousness) has the advantage of high bandwidth: it efficiently summarizes a large amount of behavioral information, and can predict a variety of relevant criteria (John, Naumann, & Soto, 2008; Soto & John, 2016). The Big Five mostly captures personality at a very abstract level, though narrower facets of personality are available. Several studies have concluded that predictions achieved by broad personality factors could be substantially increased by narrow personality traits, which suggest that the use of multiple facets of personality could provide advantages especially for explaining behavior (Makransky, Mortensen, & Glas, 2013). Other researchers advocate the use of broader personality traits, pointing toward lower reliability scores of narrow personality facets. To reach adequate reliability scales and measurements would have to be lengthened by three to six times (Makransky, Mortensen, & Glas, 2013). The third research question considers these limitations and asks, *R3: Could social media provide a more detailed assessment of personality?*

Finally, the big five are mostly being tested through self-reports which are limited by the ability and the willingness of participants to report private knowledge (Greenwald et al., 2002). This leads to the last research question, *R4: Could social media testing provide meaningful advantages over self-reports?*

4. Possibilities of Social Media for Measuring Personality Discussed in Literature

The analysis of the literature included through the systematic mapping process produced at least face valid results. Half of the papers included were published in interdisciplinary journals, emphasizing the importance of multidisciplinary perspectives and collaborations. The other papers could be assigned to the expected disciplines, namely Psychology, Computational Science and Marketing. Each paper was assigned to one or more topics. Most of the papers investigated some way of *predicting personality (13)*. Other popular topics were *social networks (4), data analysis methods (5), self-presentation (2) and marketing (2)*. Only one search result was fully excluded, the book 'Theories of Personality', due to it not being available. Since this field has already been sufficiently covered through the first literature review and was not of further importance regarding the analysis, excluding was deemed acceptable. A conference paper of an already included article was used only for literature analysis. The two studies who focused on marketing were included only on a superficial level. Regarding the sample sizes, seven studies used large data samples with more than 500 participants. Four of these Big Data samples were collected through the MyPersonality App. One publication used a Facebook application specifically designed for the study (Lönnqvist, Itkonen, Verkasalo, & Poutvaara, 2014). Another study used training data from 5'000 participants recruited through the Amazon Mechanical Turk platform (Gou, Zhou, & Yang, 2014). Finally, one study extracted data automatically from MySpace online forums (Maria Balmaceda, Schiaffino, & Godoy, 2014). The smaller sample studies used mostly data that was collected through student samples (3), specific advertisement (2), Amazon Mechanical Turk (2) or used already existing datasets (2). These results are again face valid in the sense that there are two ethical ways of collecting big social media data: either through specifically designed applications or by crowdsourcing, using online labor markets. None of the studies included collaborated directly with a social media platform provider. Most of the studies used actual social media data, though some studies relied exclusively on self-reports. There was quite a variety regarding the used platforms though most frequently, data from Facebook or Twitter was analysed. Other platforms were Youtube, StudiVZ, Foursquare, Myspace and Instagram. Two studies discussed the possibilities or suggested a method to combine social media data from multiple social media platforms (Buraya, Farseev, Filchenkov, & Chua, 2017; Skowron, Tkalčič, Ferwerda, & Schedl, 2016). All big social media data samples used in the studies included have been made publicly available, which confirms the finding from Weller and Kinder-Kurlanda (2015) that researchers are inclined to make social media data available. Finally, different types and scales of data as well as different aspects of social media platforms themselves could be identified which are thought to hold promise for measuring personality. The following section presents an overview on these possibilities.

Table 2. Overview of the main literature search

Paper*	Topic**	Type of Data	Data Size	Platform	Data collection
Schwartz (2013)	P,DA	Text	75'000	Facebook	Application
Park (2015)	P	Text	66’000	Facebook	Application
Fang (2015)	P,SN	Review	138 Samples	-	Various
Seidman (2013)	SP, SN	Self-Reports	184	-	Student Sample
Kosinski (2014)	P, SN	Behaviour, Visual,	350'000	Facebook	Application
Stopfer (2014)	SP	Perceiver Ratings	103	SchuelerVZ, StudiVZ	Advertisement
Gou (2014)	P, DA	Text	5'000	Twitter	MTurk
Chorley (2015)	P	Location	174	Foursquare	Application
Buraya (2017)	P	Relationship-Status	N/A	Twitter, Instagramm, Foursquare	Available Dataset
Farnadi (2016)	DA	Text, V-Logs,	4’250	Facebook, Youtube, Twitter	Application, MTurk, Advertising
Lima (2014)	DA	Text	20’000 tweets	Twitter	Available Datasets
Lonnqvist (2014)	P, SN	Social Network	5031	Facebook	Application
Maria (2014)	P	Text	5’002	MySpace	Automated Extraction
Skowron (2015)	P, DA	Text, Pictures	62	Twitter, Instagramm	MTurk
Zhu (2013)	P	Self-Reports	309	-	Student Sample
Stoughton (2013)	P	Self-Reports	175	-	Student Sample

** Only fully included papers are listed. **P: Predicting Personality, SN: Social Networks, DA: Data Analysis, SP: Self-Presentation .*

4.1. Possibilities of Social Media Data

Types of Social Media Data

Regarding the *type of data,* the before-mentioned studies found correlations with personality in a wide range of social media data. Table 1 provides an overview of the different types of data used in the studies. The high diversity of analysed data hints at the many possibilities that social media data could provide. Some studies used visual data, like photos or video-logs, others used social network data (e.g. number of friends, interaction between friends, transitivity of networks, cross-sex friendships), behavioural data (e.g. website preferences, 'likes'), location-based data and some studies even used multiple data sources. But the most frequently used data by far was language data. There are several reasons why language data holds particular promise in comparison to other types of data. Language data is relatively easy to compute and requires little interpretation to be useful. Visual data, e.g. pictures or videos, for example, could also be correlated with certain personality dimensions, but visual data is more challenging for computational analysis. Some studies therefore just counted the numbers of pictures on a social media platform (Kosinski, Bachrach, Kohli, Stillwell, & Graepel, 2014). A preliminary study by Skowron et al. (2016) proposes to use different aspects, like brightness, saturation, hue-related and content-based features such as a person's face or full body, as it has been done in emotion detection. Others relied on perceiver ratings for the interpretation of pictures or videos (e.g. Stopfer, Egloff, Nestler, & Back, 2014). Chorley et al. (2015) furthermore found correlations between personality and the places we visit. To date though, only little research was conducted regarding the relation between location data and personality, due to barriers in data collection (Chorley, Whitaker, & Allen, 2015). Finally, social network data, e.g. degree (number of friends), network transitivity (whether friends are friends with each other) or the sex distribution of friendships could provide unique possibilities. Maria Balmaceda, Schiaffino, & Godoy (2014) for example found that there are patterns between personality dimensions in communication threads. Agreeable people tend to communicate more often with extroverted people or people with certain personality traits, such as extraversion, agreeableness and openness to experience, tend to talk more with people with similar personality traits. It is important to note that social network data always includes some kind of interaction between different personalities. For example, every friend request needs to be accepted from another person. Social network data therefore includes behavioural data from different persons, which is influenced again by different personalities. This aspect might apply to most of the data found on social media and makes it difficult to distinguish between the influence of the user's personality and influences of the social surroundings. For example, if a certain user 'likes' a certain post on Facebook, the user usually knows very well who posted it. The 'Like' might therefore say more about the relationship between these users than about the personality. This can be seen as a limitation, which needs to be worked around or it can be seen as a chance to integrate personal and social behaviour theories. Fang et al. (2015) did a comprehensive meta-analysis, including 138 independent samples, and found that an individual's personality and his network position both are important aspects which influence job performance and career success. These findings suggest that personality and social network data found on social media platforms could provide an integrated view. Nonetheless, this points toward an important distinction regarding the scales of behavioural data. Personal and social network data need to be clearly distinguished, if interactions are to be measured. Finally, some studies hint towards the possibility of combining different cues (Skowron, Tkalčič, Ferwerda, & Schedl, 2016).

Social Media Big Data

Regarding the scales of behavioural data, social media platforms provide data on a personal scale as well as on the scale of social networks. But the most interesting scale might be the mass-behavioural scale. The MyPersonality app for example collected data from over 6 million users over years. Such high-volume, high-velocity and high-variety data is being discussed under the concept of 'Big Data' (Bello-Orgaz et al., 2016). Big data might hold the most promise and at the same times pose the biggest challenges for social media research. Big amounts of data provide problems for traditional data analysis algorithms and techniques. For this reason, the methodologies and frameworks behind the Big Data concept are becoming very popular in a wide number of research and industrial areas (Bello-Orgaz et al., 2016). Big data samples can facilitate the discovery of patterns that might not be seen in smaller samples, and the high statistical power of Big Data analyses might furthermore address the replicability crisis in psychological sciences (Kosinski et al., 2016). There are mainly two dimensions of Big data which are of particular promise. First, longitudinal data could provide meaningful possibilities and insights into personality, especially into changes of personality over time or even generations. In the literature found, little longitudinal data has actually been analysed. The MyPersonality app provides certain longitudinal data, which was for example used to measure the retest-reliability of the personality prediction approach (Park et al., 2015). The other dimension of interest is the size of the population. As stated before, personal data provides insight into personality, whereby social data provides interest into social networks and their influences. The results of whole populations furthermore could be used to gain insights into 'populations'. Psychological testing usually utilizes a certain reliable method of testing, though test scores by themselves can not be interpreted. To interpret test scores, norm-referenced tests compare results with a standardised population, showing how a certain person compares to a broader population (Rust, Golombok, Kosinski, & Stillwell, 2014). Big scales of behavioural data could therefore be interesting for measuring personality as well, simply by comparing data with a broader population. From the literature included though, Big Data samples were mostly used for data analysis purposes. Big language samples were for example being used to generate new, context adequate lexical bases, which in turn could be used to predict personality. Traditional language analysis uses 'closed vocabulary approaches', whereby Park et al. (2015) proposes to use open vocabulary approaches. Open vocabulary differ, since they create models built from words, phrases and even generate topics automatically. They seem to be useful to provide new insights which are not found by traditional closed-vocabulary approaches (e.g. that basketball correlates with emotional stability or that introverts are interested in Japanese media; Park et al., 2015). Using open vocabulary approaches might also achieve significantly higher prediction accuracies than standard lexica (Schwartz et al., 2013). Such automated vocabulary approaches could additionally be able to find counterintuitive correlation (e.g.correlations between Curly Fries and intelligence), which human raters might not find (Kosinski, Stillwell, & Graepel, 2013).

4.2 Structural Possibilities of Social Media Platforms

Some studies pointed toward possibilities which are not based on the data but on the social media platforms themselves. There are different aspects which could prove to be interesting. First, since social media platforms are online environments, there are several computational possibilities. Second, social media are of great interest because the data is generated naturally and by highly motivated people, who are willing to disclose a lot of personal information. Finally, crowdsourcing certain tasks could prove useful for researchers.

Computational Possibilities

Using the computational aspects requires again a specific skillset. With that given, there are interesting possibilities. Research has shown that certain aspects of gamification can prove useful for enhancing motivation and participation of users (Hamari, Koivisto, & Sarsa, 2014). Additionally, computerised tests could make use of adaptive testing. Further research regarding adaptive testing and personality found that personality tests can be substantially shorter without attenuating precision when adaptive testing is used (Makransky et al., 2013). Adaptive algorithms change dynamically to measure latent constructs while minimizing the number of questions each respondent must answer. The method is an extension of item response theory, whereby each question item is classified based on both its average level of 'difficulty' and its capacity to discriminate between responses (Montgomery & Cutler, 2013). Adaptive testing is suggested to be particularly appropriate for constructs that have many highly-correlated facets (Makransky et al., 2013).

Motivational possibilities

Motivational aspects are important regarding user participation and endurance. Especially for collecting longitudinal data, participants are required to generate data over a long period, where such motivation might be difficult to uphold. The traditional way to motivate partcipants is through specific Incentives, e.g. raffles or other financial incentives, or credits for student samples. Interestingly though, many studies only provided feedback on the personality of participants as an incentive, which was found to be sufficient. The MyPersonality app for example did not provide any financial incentives and collected data from over 6 million users. Regarding the motivation to use social media platforms, self-representation, maintaining social networks and enjoyment have been suggested as some of the most important aspects (Lin & Lu, 2011). The need to self-represent is of particular interest regarding measuring personality. Since people tend to self-represent accurately (Back et al., 2010), this major motivation of social media users does align very well with the interests of personality researchers. Both essentially want accurate presentations of the users personality. Self-presentation it thought to be accomplished for example by posting photographs, profile information and wall content (Seidman, 2013). Many more ways can be imagined, how users can present themselves online (e.g. by posting feedback from personality measures). More theoretically, self-presentation is thought to be achieved through impression management, the process by which individuals control the impressions others form of them (Stopfer et al., 2014). Impression management is closely related to the social desirability bias, one of the main limitations of self-report measures. There is ample empirical evidence that respondents systematically overreport socially desirable behaviors and attitudes and systematically underreport socially undesirable traits (Krumpal, 2013). Additional literature research regarding impression management and social desirability has been conducted and found that there are two main factors or theoretical explanations that are being discussed for social desirability biases: self-deception and impression management (Blasberg, Rogers, & Paulhus, 2014). Regarding impression management, respondents strive for social approval via selecting the answer that is expected to lead to positive social evaluations and minimizes negative reactions. In contrast, the concept of self-deception assumes that interviewes want to maintain a positive self-image to maximize self-worth and to reduce cognitive dissonance resulting from divergence between social norms, self-perception and reality (Krumpal, 2013). Impression management reflects the subject's actual behaviour based on what they think is right and good in the context of their relations with others and society in general (Elliot et al., 2016). Desirable answers can also be conceptualized as respondents' temporary social strategies to cope with different situational factors in surveys (Krumpal, 2013). Regarding social desirability biases, different approaches to reducing effects of faking have been explored (Stark et al., 2014). Though certain aspects, for example anonymity, are known to limit social desirability bias, no significant differences between online, offline and paper surveys was found regarding social desirability bias (Dodou & de Winter, 2014). This studiy did account for the potential that social media platforms could provide for impression management. Since real social networks are a structural and defining aspect of social media platforms, this might provide new possibilities. Gou et al., 2014 found that participants were generally inclined to share their personal information even at work, but wanted to be in control of what is being shared, with whom, when and how often. Especially since social media platforms are increasingly being searched by third parties, for example potential future employers, the need for impression management is very understandable. Social media platforms provide the possibility to implement impression management as an inherent part of testing, for example by providing participants to report only certain data or even fake results towards their peers. Including impression management into a social media testing platform could therefore enable people to report their true selves, while still maintaining control over the impression that others form. Further research in this direction could prove very interesting. Finally, the marketing literature found during the literature research focussed on the possibilities how social network users could get actively engaged in certain social media groups or applications (Ngai, Moon, Lam, Chin, & Tao, 2015; Walsh, Clavio, Lovell, & Blaszka, 2013). Lessons learned from marketing could be interesting with respect to further enhance motivation or to target specific groups which otherwise might be underrepresented. This could be used to further increase the diversity and representativeness of the population.

Possibilities of Crowdsourcing

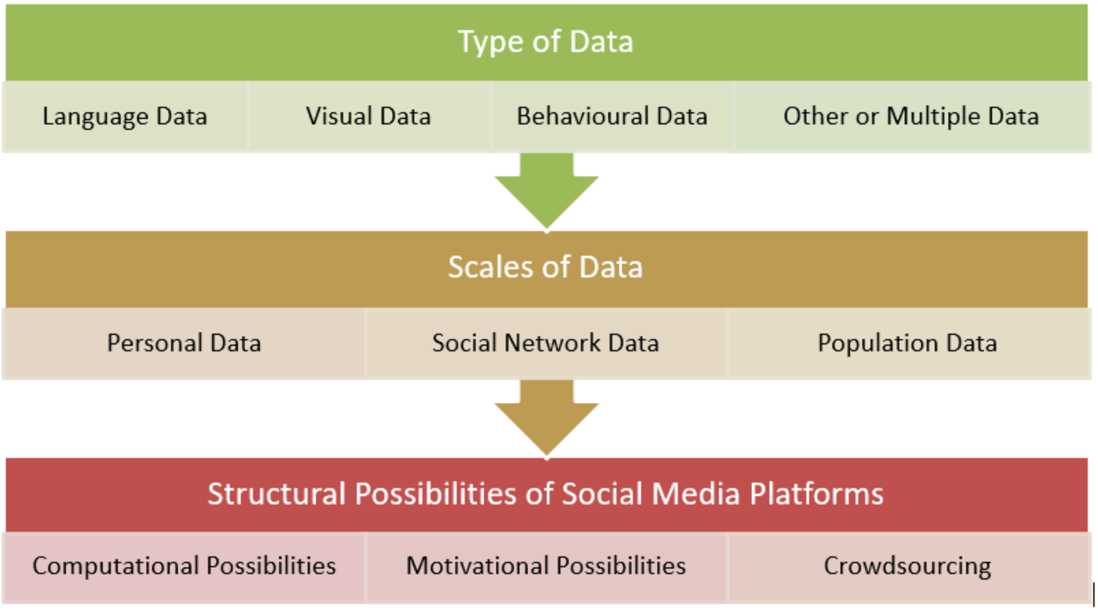
Finally, 'crowdsourcing' could provide additional possibilities. Crowdsourcing can be defined as a type of participative online activity in which a crowdsourcer proposes to a diverse group of individuals to undertake a specific task. The undertaking of the task entails mutual benefit, whereby the user will receive some sort of incentive, be it economical, social recognition or feedback, while the crowdsourcer will obtain the results (Estellés-Arolas & González-Ladrón-De-Guevara, 2012).

Crowdsourcing is not limited to applications like Amazon Mechanical Turk, but is rather an inherent part of social media platforms. As stated before, financial incentive does not seem to be required to recruit social media users for research purposes as long as feedback about their personality is provided. This motivation might enable researchers to outsource other specific tasks to social media users as well if motivation is being ensured. Finally, further research in to the question whether the collective intelligence of crowds is superior to individual intelligence for specific tasks could provide interesting applications (Kosinski, Bachrach, Kasneci, Van-Gael, & Graepel, 2012).

Conceptual Framework of the Possibilities of Social Media Platforms

The data so far showed many interesting opportunities social media platforms could present for measuring personality. To create a comprehensive overview of these possibilities a conceptual framework was deemed useful. To my knowledge, no such framework does yet exist. The closest found is the one by Ngai, Moon, Lam, Chin, and Tao (2015) who provided a conceptual framework for social media application development. This review focused on a marketing perspective, but the theoretical framework proved nonetheless to be useful. Certain adaptions had to be made though. Ngai et al. (2015) distinguish between personal theories, social theories and mass-communication theories. As shown before, this distinction is quite useful regarding the possibilities of social media platforms: personal behaviour data, social network data and mass-communication data (or Big Data) provide different possibilities and should therefore be distinguished. Careful distinctions are especially required between social influences and personal influences, since those are regularly intertwined at the level of social media platforms. Ngai et al. (2015) further distinguish between different social media tools and technologies, e.g. blogs/microblogs (Twitter) or social networking sites (Facebook, LinkedIn). Which platform was used seemed to be of limited importance, especially since the different platforms tend to be increasingly connected or even combined, blurring out clear distinctions. The terminological distinction proposed by Ngai et al. (2015) and others between different social media technologies was therefore not adopted for this study. But the distinction still proved useful, since it pointed towards the possibilities that lie within social media platforms, e.g. computational, motivational or crowdsourcing possibilities. The conceptual framework of the possibilities of social media platforms presented in figure 1 is thought to be useful especially regarding the development of new methods, platforms or applications, with the goal of measuring personality.

Figure 1: Theoretical framework of the possibilities social media platforms could provide



5. New Ways for Measuring Personality

The following section tries to answer the research questions that have been worked out in the initial literature research. The answesr are based on the information contained in the scope of this small sample scoping review. Since the scope of this review is very limited and might not encompass important literature, the following answers should be interpreted with caution.

R1: Could social media provide new ways for measuring the within-person variance?

There are two main aspects identified regarding the possibilities of measuring the within-person variability. First, social media could provide longitudinal data which might enable studying within-person variations over time. Social media platforms include data on millions of people over, presumably, many years to come. Such data could provide sources for understanding not only if personality changes, but how personality changes as well. Schwartz (2013) for example provides a visualization of how language use changes at different age groups showing that family-related terms are much more frequently used by people over thirty years. Though such evidence is only face-valid and there are different biases to consider (e.g. cultural biases for certain language use), longitudinal data still holds a lot of promise, especially if collected on a grand scale. It could provide insight into within-person variations during lifetime and even generations. It could also be used to research the influences of specific incidents on personality, for example by comparing the personality scores before and after a specific incident (e.g. after a catastrophe). Collecting big personal data over a long period presents challenges for the motivation of participants and regarding ethics. From an ethical perspective, it seems very questionable to rely on commercial social media platforms for such data. Researchers therefore might need independend ways to collecting data. Even if found, who could be trusted with such data? Mittelstadt and Floridi (2016) suggest a clear distinction between commercial and academic uses, which is important, but even trusting such data to a specific academic institution seems problematic. An open-science project, enabling access to researchers from all around the world, which secures anonymity and clearly communicates how and for what scientific purposes the data is being used, might be the most adequate ethical solution. If researchers want to collect the data independent, it seems important to consider the motivational aspects which are required to keep users providing longitudinal data. A scientific application to measure personality would need to ensure motivation of users, for example through elements of gamification, strategic marceting and by providing feedback about the test results. Even without using longitudinal data, situational variables found on social media platforms could provide important insights into the situational influences on the within-person variability. Of particular interest are social influences, which are ubiquitous on social media platforms. It seems important though to note, that to understand the different influences they should be measured independently. Much of the social media data available requires some form of interaction between users and might therefore depend on more than one personality as well as situational factors. It therefore might be helpful to control the online environment in which social media data is collected and for example limit or isolate social influences. Other situational variables are locations and behaviours at specific times, which also could be of interest (e.g. how does personality vary after someone visited a restaurant). Many more situational variables could be collected directly (e.g. What website was visited before a specific behaviour occured) or through self-reports. Furthermore, datasets could be combined with crowdsourced self-reports which would allow to asses even more situational variables, e.g. velocity. These findings suggest, that social media platforms provide many new and exiting possibilities for measuring and understanding the within-person variance of personality and could provide an important step towards further integrating situational and personal variables in to a more complex understanding of personality.

R2: Could social media provide new ways regarding the lexical base?

The Big Five do not try to account for the full complexity of individual behaviour but rather focus on traits (Matz, Chan, & Kosinski, 2016). The lexical base used for the Big Five therefore consists only of words describing traits (Matz, Chan, & Kosinski, 2016). Social media data includes all kinds of language data, including for example swear-words and emoticons as description of traits, states or social evaluations (Schwartz et al. 2013). Language-based assessments using open-vocabulary approaches can generate new lexical bases out of Big Data sets, which can then be used to predict personality (e.g. Park et al., 2015). To my knowledge, it has not been tested whether such lexical bases could be used to replicate the Big Five or even generate new dimensions of personality. To do so, either a more complex concept of personality would be required (which is able to include more than just traits) or the language data would have to be limited. Assuming that all language data could be used to distinguish invidual behaviour, it might even be possible to find more complex concepts of personality using factor analysis of more diverse language data. Open vocabulary approaches have already been used to automatically generate different topics (Park et al., 2015). Whether all this data could really be included into one theoretical concept is very questionable though, since no such all-encompassing theory of personality or behaviour exists yet (Rust, Golombok, Kosinski, & Stillwell, 2014). There is a theoretical concept though which integrates different dimensions already, the integrated trait-state model (Hamaker, Nesselroade, & Molenaar, 2007). Using situational and personal data might be useful to further develop the integrated trait-state model of personality. Using only specific classes of language data could maybe even be used to generate more complex multidimensional concepts of personality which combine for example traits, states and social influences. A different approach which would not require a new concept of personality, but could still make use of all the available data could be using mass-behavioural data. Big enough data sample could be used as a standard population which in turn would allow to assess all kinds of individual language behaviour in comparison to the standard population as it is known from norm-referenced tests (Rust, Golombok, Kosinski, & Stillwell, 2014). Such results might be very difficult to interpret (e.g. what does it mean, if some individual uses double the amounts of emoticons than an average user?) but such an approach might still be useful for certain research and to provide additional insights (e.g. how is emoticon use distributed?). If not all language data can be conceptualized, language data would have to be limited for certain classes or types of word, e.g. traits or states (or both). This would require to limit the social media language data either at the point of measurement (data input) or at the stage of data analysis. Social media platforms themselves mostly do not limit the language (exept e.g. racial slurs). Researches could manipulate data input by using social media applications or other scientifically designed online environments which would allow, for example, only specific language data input. Another approach would be to identify the relevant language data in the mass of the broader language data available. Computational factor analysis could be helpful and even find counter-intuitive correlations (Kosinski 2013). Future research is required but seems very promising. A different argument against the lexical bases used in the Big Five are biases, which might be inherent to language data (Trofimova, 2014). Language is said to be a social construct and not actual behaviour, therefore the lexical hypothesis could be invalid (Trofimova, 2014). This critique certainly does not apply to social media language data. Trofimova (2014) argues, that the lexical approach is an analysis of relationship between lexical description of behaviour but not actual behaviour. Language data collected through social media displays actual language behaviour and not just descriptions of it and could therefore provide a significantly better lexical base. Furthermore, social media language data could stem from a more diverse, multicultural and multilingual population (Kosinski, Bachrach, Kohli, Stillwell, & Graepel, 2014) and the internet might be a surrounding, which is less influenced by pro-social behaviour, due to its anonymity and the distance to other users (Dodou & de Winter, 2014). Social media data might be influenced by several other biases though (Morstatter, Pfeffer, & Liu, 2014). To control and adapt accordingly to such biases, control over the data collection process is required.

R3: Could Social media provide a more detailed assessment of personality?

There are certain studies who go beyond the big five and invite other personality models as well (e.g (Gou, Zhou, & Yang, 2014). This indicates that the predictive power might not be limited to abstract dimensions of personality. Park et al. (2015) researched correlations between language based assessments and self-report on the narrower facet level and on the broader trait level, finding correlations for both levels. This implicates that language based assessments could be used as well to provide assessments of narrower facets. Language based assessments do not really require additional data and seem to be able to compute facet level prediction with the same dataset than broader dimensions (but maybe not as reliable; Park et al., 2015). Questionnaires could use adaptive testing to shorten questionnaires (Makransky et al., 2013), which in turn could allow better assessment of facet level personality without lengthening tests. Furthermore, some specific types of data could prove to be specifically relevant for certain facets, e.g. certain Website-Choices and 'Artistic Interests' (as a facet of Openness). More detailed assessments of personality might require again a more complex and multidimensional model of personality. Narrower facets, for example 'altruisms' as a facet of agreeableness, might depend heavily on social or other situational influences. A concept of personality which could account for stable traits as well as social influences, might be able to predict narrower traits more reliable. A multidimensional approach might also enable to locate narrower facets on different dimensions more clearly. Finally, narrower facets are thought to provide advantages especially for explaining behavior (Makransky, Mortensen, & Glas, 2013). This suggests, that narrower facets would correlate higher with actual behaviour than the big five, a suggestion that could not be confirmed with the literature included (Park et al., 2015). Still, if narrower facets correlate higher with actual behaviour than with more abstract test scores, the wide amount of behavioural data could in turn be very useful for predicting narrower facets.

R4: Could Social media testing provide meaningful advantages over self-reports?

There are several papers hinting towards possibilities social media testing could provide over self-reports. Language based assessments for example are relatively fast and cheap and agree with self-report questionnaires (Park et al., 2015). They can be easily shared and adapted as computer code as well and could complement and extend traditional measures of self-reports in social media samples already by providing an alternative (Park et al., 2015). A further advantage is that they can be generated retroactively, giving researchers an alternative method to study past behavior without relying on participants' memories (Park et al., 2015). Social media platforms could also provide new possibilities, where self-reports traditionally lack, specifically regarding the willingness to report private knowledge. The willingness of participants to provide private knowledge can be understood, amongst other things, as a social desirability bias. Social desirability is at least partially intentional, where participants strive to control the impression others form of them. Social media platforms provide unique possibilities regarding such impression management, since social networks are an integral part of social media platforms. It could therefore be very interesting to include impression management and, for example, enable participants to specifically control what information is being shared with whom. In a specifically designed online environment to measure personality, it would be possible, for example, to clearly distinguish between the results presented to the users and the results presented to the social network, even enabling participants to actively manipulate the results their friends and acquaintances see to their benefit. This could fulfil the need to manage one's impression while at the same time providing researchers with true data.

6. Experiences and Limitations of the Chosen Method

There are several important limitations to this study. First, the scope of the literature included was very limited. The intuitive literature review which was conducted at the beginning of the research accounted for a broader scope, but this *small sample* scoping review must be clearly distinguished from systematic literature or scoping reviews, which are able to picture the research landscape adequately and comprehensively. The chosen approach, to limit analysis for the first few results, is furthermore questionable. Two main objectives were pursued with this approach: creating an overview over the research landscape and counteracting the confirmation bias by providing some sort of objective inclusion criteria. As stated before, whether such an approach is useful for picturing the research landscape depends mostly on the ranking algorithms used by Google Scholar, which are not disclosed. The results were at least face-valid which indicates that the literature included with this approach was in fact relevant. Whether the results are representative can not be answered without additional research. The initial research though seems to have included, for example, more critical evidence than the first twenty search results, which points toward a publication bias. The sample is therefore thought to be not representative for the broader research landscape, but still useful for providing an idea of the research landscape and informing future research. The search queries used were furthermore not developed by a specialist and might have been imprecise. The main search query (‘social media’ OR ‘social networks’ AND personality) was intended to include ‘Social Online Networks’ and ‘Social Networking Sites’, which are synonyms often used for social media platforms. Additional search queries with more precise search terms (e.g. ‘Social Media’ OR ‘Social Networking Sites’ OR ‘Social Online Networks’ AND ‘Personality’) were conducted and resulted, at least for the first 10 results, in very similar results (9/10 were identical but only 16/20). This indicates that precise search terms are of increasing importance, as the number of papers included rises. The second goal, counteracting the confirmation bias by using some sort of objective inclusion criteria is thought to be reached through this approach. This approach led to the inclusion of, at first glance, seemingly irrelevant literature and therefore expanded the understanding. It should be pointed out though that this objectivity might be at the cost of a publication bias towards including only ‘popular’ papers. For the goal of gaining an overview of the possibilities discussed in recent literature, this popularity bias though seems acceptable, since it is thought to be a good indicator of research areas that receive a lot of attention and therefore are thought to hold a lot of potential. Finally, a limitation of this study might be seen in the deviation from the original theoretical framework. The mixture of intuitive literature research and systematic research was not intended in the original framework by Arksey. Due to the specific situation in which this research was conducted, these deviations though seemed adequate. Especially regarding the creative task of finding relevant topics and possible research gaps, the intuitive literature research was found to be helpful. In contrary, the systematic research was thought to limit the confirmation bias and was useful for analysing the literature. Since the literature accounted for in the intuitive literature review is not clearly identifiable, this approach does limit the replicability which other forms of systematic review hold. This limitation is only partial and in regards to literature which was scanned but not included in this final version. This review did not assess the results or the quality of the included papers, mostly because listing results without assessing the quality could be misleading and assessing the quality is not required in scoping reviews (Arksey & O’Malley, 2005). In summary, this small sample scoping review was found useful for providing an idea of a research landscape, create a theoretical framework and inform future research. The mixed method allowed creative research as well as some form of objective inclusion criteria, at the partial cost of replicability.

The small sample size is not thought to be representative of the broader research landscape but to provide an idea over relevant and highly cited literature in this area. It should be clearly distinguished from broader scoping reviews, which are able to picture research landscapes adequately. At the same time though it should be distinguished from traditional unsystematic literature research, due to its ability to provide analysis of the literature included and the higher objectivity in comparison to traditional literature research. Further research could use different scopes (for example the first 300 papers) and/or different search engines for different research questions and therefore consider different balances between breadth and comprehensiveness and practical issues related to time, funding, and access to resources (Levac et al., 2010). Additional experiences are required, whether such an ‘objective’ limitation, as seen in the number of search results included, could be useful regarding the lack of clear boundaries, increasing feasibility (O’Brien et al., 2016). Small sample scoping studies could prove useful in situations, where time is very limited, for example for side projects or in the context of university assessments with fixed deadlines. Since the scope of the studies included can be varied according to the specific circumstances, this approach could be an interesting and flexible addition to the systematic literature review methodology.

7. Conclusion

Social media platforms do provide many different interesting possibilities for measuring personality. Especially the possibilities that Big Data provides are of great interest and could alter not only the measurement but also our understanding of personality. Big data samples allow new ways to predict personality, based on a wide variety of data types, and could provide a standard population with which an individuals behaviour could be compared. Lexical bases including actual language behaviour of millions of people could furthermore, theoretically, be used to generate multidimensional concepts of personality, which account for more dimensions of personality than just stable traits. The overall purpose of this review though was to determine the value of a multidisciplinary collaboration for developing an online platform or social media application to measure personality. Literature and research included in this sample seems to focus mostly on the different possibilities to analyse and compute the available data. The collection of this data seems to provoke little interest. To collect social media data, two ethical approaches could be identified: a) collecting social media data or self-reports through specific applications and b) crowdsourcing specific tasks through online labor markets. Specific applications do not require financial incentives and can collect data from up to millions of users (e.g. MyPersonality App). They do require specific skills to compute and need to be appealing to users. Crowdsourcing on the other hand provides the possibility of oursourcing very specific tasks (which participants otherwise might not execute with the necessary care) and enables precise control over the data. These approaches seem to differ mainly regarding the participant’s motivation and incentives. Social media applications don’t require financial incentives and provide the participant’s only with feedback about their personality. This might be understood best, considering that self-presentation is one of the main reasons why people use social media platforms (Lin & Lu, 2011). Self-presentation though should be understood as a challenging process for social media users. It requires insights into one’s true self to adequately self-present and should simultaneously fulfil the need to manage the impression others form of oneself. Regarding the insight into one's true self, interest of users and personality researches are very much aligned. The need to manage the impression of others might not be of direct interest to personality researchers, but integrating this need in the collection of personal data could limit social desirability biases and increase the willingness of users to report private knowledge. These motivational resources are suggested to be some of the most important possibilities that social media platforms provide for measuring and understanding personality and a necessary condition for all further applications. People from all around the world, want to know who they are and want to present their actual selves online. Helping those people to present themselves adequately could in turn allow researchers to gain invaluable insights into individual, social and mass-scales of different kinds of behaviour. Optimizing this motivation through marketing, gamification and non-financial incentives could therefore enable longtime collection of all kinds of big personal data for scientific purposes. Research is thereby neither limited nor depending on 'normal' social media data. Instead, researchers should aim to control the environment in which data is collected by computing specific social media applications or other scientific online environments. This independence from commercial social media sites is important out of ethical reasons and provides additional possibilities regarding the quality and control of the data. It could, for example, be possible to create an application, which allows only specific types of input, e.g. language data which describes traits or states. This control over the data collection might also enable to isolate, for example, social influences from personal behaviour, which could provide important insights into the interactions between personality and social networks. Such an application does indeed require a multidisciplinary collaboration from at least psychology, computational science and marketing. Out of ethical and practical reasons, it should be done as an open-science research project, inviting all researchers who are interested and willing to contribute. This review suggests, that such a collaboration could be of great value and provide many new possibilities for measuring and understanding personality.

Acknowledgement

The author would like to thank Dr. Alexander Weiss and Jannis Vamvas for their suggestions and feedback regarding an earlier form of this manuscript.

Declaration of scientific integrity

The author hereby declares that he has read and fully adhered the Code for Good Practice in Research of the University of Basel.

References

Allport, G. W. (1937). *Personality*. Holt New York.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8* (1), 19–32.

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The hexaco honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, *18* (2), 139–152.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*, *21* (3), 372–374.

Badenes, H., Bengualid, M. N., Chen, J., Gou, L., Haber, E., Mahmud, J., . . . others (2014). System u: automatically deriving personality traits from social media for people recommendation. In *Proceedings of the 8th acm conference on recommender systems* (pp. 373–374).

Beel, J., & Gipp, B. (2009). Google scholar's ranking algorithm: an introductory overview. In *Proceedings of the 12th international conference on scientometrics and informetrics (issi'09)* (Vol. 1, pp. 230–241).

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social Big Data: Recent achievements and new challenges. *Information Fusion*, *28* , 45–59.

Blasberg, S. A., Rogers, K. H., & Paulhus, D. L. (2014). The bidimensional impression management index (bimi): Measuring agentic and communal forms of impression management. *Journal of personality assessment*, *96* (5), 523–531.

Borkenau, P., & Ostendorf, F. (1998). The big five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality*, *32* (2), 202–221.

Buraya, K., Farseev, A., Filchenkov, A., & Chua, T.-S. (2017). Towards user personality profiling from multiple social networks. In *Proceedings of the thirty-first aaai conference on artificial intelligence. aaai.*

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2016). Amazon mechanical turk in organizational psychology: An evaluation and practical recommendations.

Journal of Business and Psychology, 1–15.

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

Conway, M., & O'Connor, D. (2016). Social media, Big Data, and mental health: current advances and ethical implications. *Current opinion in psychology*, *9* , 77–82.

Daudt, H. M., van Mossel, C., & Scott, S. J. (2013). Enhancing the scoping study methodology: a large, inter-professional team's experience with arksey and o'malley's framework. *BMC medical research methodology*, *13* (1), 48.

Davis, K., Drey, N., & Gould, D. (2009). What are scoping studies? a review of the nursing literature. *International journal of nursing studies*, *46* (10), 1386–1400.

Dodou, D., & de Winter, J. C. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, *36* , 487–495.

Elliot, A. J., Aldhobaiban, N., Murayama, K., Kobeisy, A., Gocłowska, M. A., & Khyat, A. (2016). Impression management and achievement motivation: investigating substantive links. *International Journal of Psychology*.

Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, *38* (2), 189–200.

Fang, R., Landis, B., Zhang, Z., Anderson, M. H., Shaw, J. D., & Kilduff, M. (2015). Integrating personality and social networks: A meta-analysis of personality, network position, and work outcomes in organizations. *Organization Science*, *26* (4), 1243–1260.

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., Davalos, S., Moens, M., De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, *26* (2-3), 109–142.

Gou, L., Zhou, M. X., & Yang, H. (2014). Knowme and shareme: understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 955–964).

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review*, *109* (1), 3.

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. (2007). The integrated trait–state model. *Journal of Research in Personality*, *41* (2), 295–315.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?–a literature review of empirical studies on gamification. In *System sciences (hicss), 2014 47th hawaii international conference on* (pp. 3025–3034).

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*(2-3), 111-135.

Huang, T., Lan, L., Fang, X., An, P., Min, J., & Wang, F. (2015). Promises and challenges of Big Data computing in health sciences. *Big Data Research*, *2* (1), 2–11.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3* , 114–158.

Keele, S. (2007). Guidelines for performing systematic literature reviews in software engineering. In *Technical report, ver. 2.3 ebse technical report. ebse.* sn.

Kosinski, M., Bachrach, Y., Kasneci, G., Van-Gael, J., & Graepel, T. (2012). Crowd iq: Measuring the intelligence of crowdsourcing platforms. In *Proceedings of the 4th annual acm web science conference* (pp. 151–160).

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, *95* (3), 357–380.

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70* (6), 543.

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.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining Big Data to extract patterns and predict real-life outcomes. *Psychological Methods, 21 (4), 493.*

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, *47* (4), 2025–2047.

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation Science*, *5* (1), 69.

Lima, A. C. E., & De Castro, L. N. (2014). A multi-label, semi-supervised classification approach applied to personality prediction in social media. *Neural Networks*, *58* , 122–130.

Lin, K.-Y., & Lu, H.-P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, *27* (3), 1152–1161.

Lönnqvist, J.-E., Itkonen, J. V., Verkasalo, M., & Poutvaara, P. (2014). The five-factor model of personality and degree and transitivity of facebook social networks. *Journal of Research in Personality*, *50* , 98–101.

Lucas, R. E., & Donnellan, M. B. (2009). If the person–situation debate is really over, why does it still generate so much negative affect? *Journal of Research in Personality*, *2* (43), 146–149.

Makransky, G., Mortensen, E. L., & Glas, C. A. (2013). Improving personality facet scores with multidimensional computer adaptive testing: An illustration with the neo pi-r. *Assessment*, *20* (1), 3–13.

Maria Balmaceda, J., Schiaffino, S., & Godoy, D. (2014). How do personality traits affect communication among users in online social networks? *Online Information Review*, *38* (1), 136–153.

Martin-Martin, A., Orduna-Malea, E., Harzing, A.-W., & López-Cózar, E. D. (2017). Can we use google scholar to identify highly-cited documents? *Journal of Informetrics*, *11* (1), 152–163.

Matz, S., Chan, Y. W. F., & Kosinski, M. (2016). Models of personality. In *Emotions and personality in personalized services* (pp. 35–54). Springer.

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of Big Data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, *22* (2), 303–341.

Montgomery, J. M., & Cutler, J. (2013). Computerized adaptive testing for public opinion surveys. *Political Analysis*, 172–192.

Ngai, E. W., Moon, K.-l. K., Lam, S. S., Chin, E. S., & Tao, S. S. (2015). Social media models, technologies, and applications: an academic review and case study. *Industrial Management & Data Systems*, *115* (5), 769–802.

O'Brien, K. K., Colquhoun, H., Levac, D., Baxter, L., Tricco, A. C., Straus, S., Wickerson, L. Nayar, A., Moher, D., O'Malley, L. (2016). Advancing scoping study methodology: a web-based survey and consultation of perceptions on terminology, definition and methodological steps. *BMC Health Services Research*, *16* (1), 305.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108* (6), 934.

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research synthesis methods*, *5* (4), 371–385.

Rust, J., Golombok, S., Kosinski, M., & Stillwell, D. (2014). *Modern psychometrics: The science of psychological assessment*. Routledge.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., . . . others (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, *8* (9), e73791.

Seidman, G. (2013). Self-presentation and belonging on facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, *54* (3), 402–407.

Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016). Fusing social media cues: personality prediction from twitter and instagram. In *Proceedings of the 25th international conference companion on world wide web* (pp. 107–108).

Soto, C. J., & John, O. P. (2016). The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power.

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From able to tapas: A new generation of personality tests to support military selection and classification decisions. *Military Psychology*, *26* (3), 153.

Stillwell, D. J., & Kosinski, M. (2004). mypersonality project: Example of successful utilization of online social networks for large-scale social research. *American Psychologist*, *59* (2), 93–104.

Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management and meta-accuracy. *European Journal of Personality*, *28* (1), 73–94.

Stoughton, J. W., Thompson, L. F., & Meade, A. W. (2013). Big five personality traits reflected in job applicants' social media postings. *Cyberpsychology, Behavior, and Social Networking*, *16* (11), 800–805.

Trofimova, I. (2014). Observer bias: an interaction of temperament traits with biases in the semantic perception of lexical material. *PloS one*, *9* (1), e85677.

Walsh, P., Clavio, G., Lovell, M. D., & Blaszka, M. (2013). Differences in event brand personality between social media users and non-users. *Sport Marketing Quarterly*, *22* (4), 214.

Weller, K., & Kinder-Kurlanda, K. E. (2015). Uncovering the challenges in collection, sharing and documentation: The hidden data of social media research. *Standards and Practices in Large-Scale Social Media Research: Papers from the*, 28–37.

Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of facebook research in the social sciences. *Perspectives on psychological science*, *7* (3), 203–220.

Wilson, R. E., Thompson, R. J., & Vazire, S. (2016). Are fluctuations in personality states more than fluctuations in affect? *Journal of Research in Personality*.

Wilt, J. A., Bleidorn, W., & Revelle, W. (2016). Velocity explains the links between personality states and affect. *Journal of Research in Personality*.

Zhu, X., Woo, S. E., Porter, C., & Brzezinski, M. (2013). Pathways to happiness: From personality to social networks and perceived support. *Social networks*, *35* (3), 382–393.

Appendix A: Research questions, relevant topics and search queries

Research Question	**A: What are the limitations of the predominant way of measuring Personality?**
Relevant Topics and Research Questions	Social Media/Social Networks + Predicting Personality, Online personality, Peer-review, Motivation, Population, Behaviour, Impression-Management, social desirability, Computational possibilities, Adaptive Testing, Language, Big Data, Collective Intelligence, Crowdsourcing, Ethics, Arksey 2005, Systematic mapping, Scoping Studys **Person-Situation Debate / Traits and States** *R1: Could Social media platforms provide new ways for measuring relevant aspects of the within-person variability?* **Lexical approach** *R2: Could Social media provide new ways regarding the lexical base of personality models?* **Bandwith-Fidelity / Abstraction-Narrowness** *R3: Could Social media provide a more detailed assessment of personality.* **Self-Reports** *R4: Could Social media testing provide meaningful advantages over self-reports* **Systematic mapping** *R5: What are the experiences and limitations using the framework of scoping review for a small sample preliminary study with the given limitations?*
Research Question	**B: What are the possibilities and limitations that social media could provide for measuring personality discussed in literature?**
Final Search Querys and Criterias	‘Social Media’ OR ‘Social Networks’ personality 20, R* Impression Management Social Desirability 5, R Personality adaptive testing 5, R Social media Big Data 10, R Arksey 2005 scoping 3, All time

* The number implies the amount of results included. R stands for recent (since 2013).